Data Security and Privacy in Distributed Collaborative Scenarios

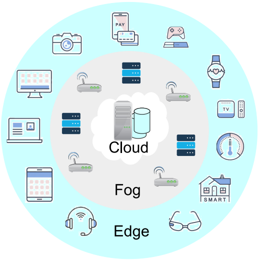
Pierangela Samarati
Dipartimento di Informatica
Università degli Studi di Milano
pierangela.samarati@unimi.it

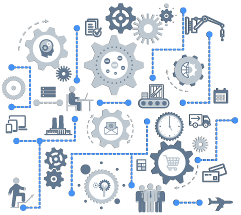
European Network for Cybersecurity (NeCS) PhD School

Cortina d'Ampezzo, Italy – January 8, 2024

ICT ecosystem

- Advancements in the ICT and networks have changed our society
- 5G and 6G, infrastructures and services are more powerful, efficient, and complex





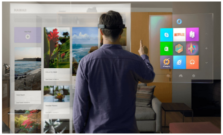
- ICT and network advancements are enabling factors for a smart society . . .

. . . Everything is getting smart

Smart car

Museum and exhibitions

Health Care


Augmented reality

Smart e-commerce

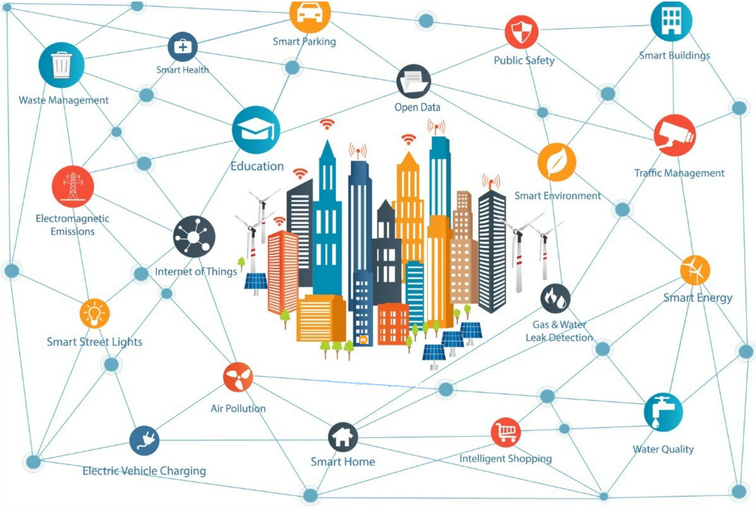
Intelligent shops

Smart entertainment systems

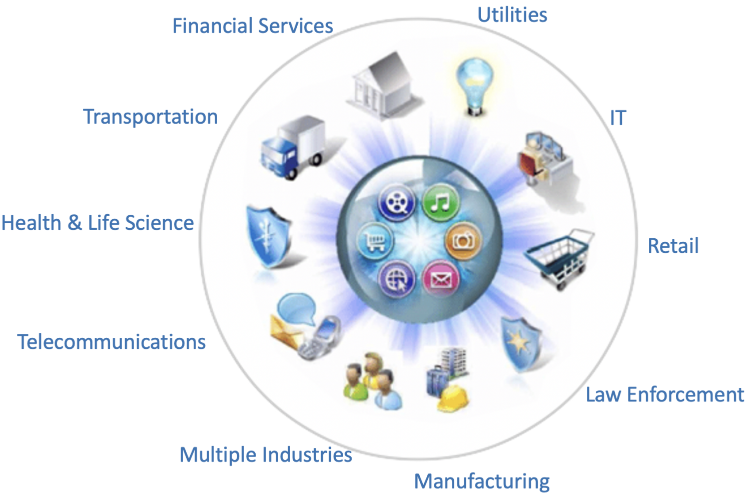
Smart governance

Smart toothbrush

Smart society



Smart society - Advantages



Smart services and security – Advantages

- + Better protection mechanisms
- + Business continuity and disaster recovery
- + Prevention and response

. . . but . . .

Smart services and security – Disadvantages

- More complexity …

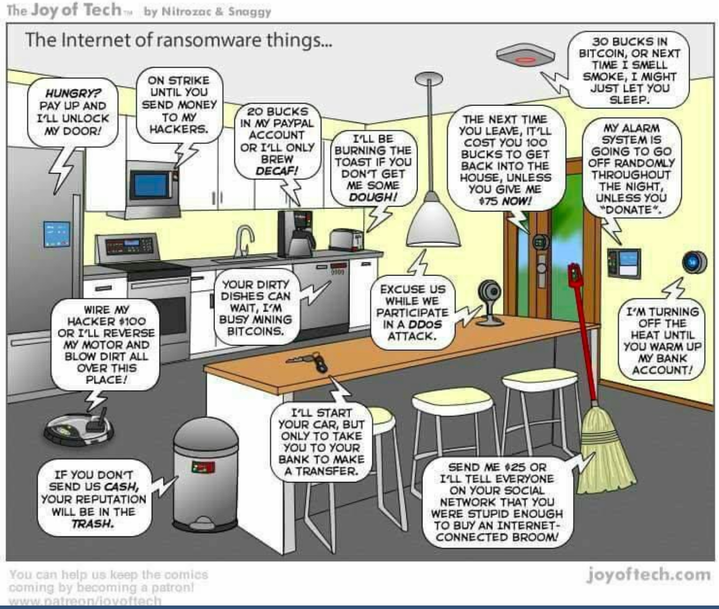
  … weakest link becomes a point of attack

  - system hacking
  - improper information leakage
  - data and process tampering

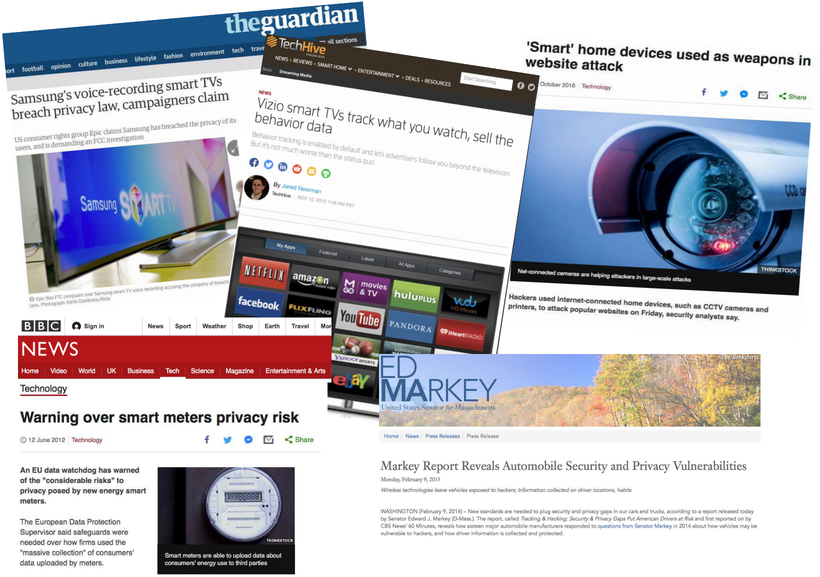
- Explosion of damages and violations

- Loss of control over data and processes

Maybe too smart? – 1


The Joy of Tech™ by Nitrozac & Snaggy
The Internet of ransomware things...
HUNGRY? PAY UP AND I'LL UNLOCK MY DOOR!
ON STRIKE UNTIL YOU SEND MONEY TO MY HACKERS.
20 BUCKS IN MY PAYPAL ACCOUNT OR I'LL ONLY BREW DECAF!
I'LL BE BURNING THE TOAST IF YOU DON'T GET ME SOME DOUGH!
THE NEXT TIME YOU LEAVE, IT'LL COST YOU 100 BUCKS TO GET BACK INTO THE HOUSE, UNLESS YOU GIVE ME $75 NOW!
30 BUCKS IN BITCOIN, OR NEXT TIME I SMELL SMOKE, I MIGHT JUST LET YOU SLEEP.
MY ALARM SYSTEM IS GOING TO GO OFF RANDOMLY THROUGHOUT THE NIGHT, UNLESS YOU "DONATE".
YOUR DIRTY DISHES CAN WAIT, I'M BUSY MINING BITCOINS.
EXCUSE US WHILE WE PARTICIPATE IN A DDoS ATTACK.
I'M TURNING OFF THE HEAT UNTIL YOU WARM UP MY BANK ACCOUNT!
WIRE MY HACKER $100 OR I'LL REVERSE MY MOTOR AND BLOW DIRT ALL OVER THIS PLACE!
I'LL START YOUR CAR, BUT ONLY TO TAKE YOU TO YOUR BANK TO MAKE A TRANSFER.
IF YOU DON'T SEND US CASH, YOUR REPUTATION WILL BE IN THE TRASH.
SEND ME $25 OR I'LL TELL EVERYONE ON YOUR SOCIAL NETWORK THAT YOU WERE STUPID ENOUGH TO BUY AN INTERNET-CONNECTED BROOM!
You can help us keep the comics coming by becoming a patron! www.patreon/joyoftech
joyoftech.com

Maybe too smart? – 2



theguardian

ort football opinion culture business lifestyle fashion environment tech trave

Samsung's voice-recording smart TVs breach privacy law, campaigners claim

US consumer rights group Epic claims Samsung has breached the privacy of its users, and is demanding an FCC investigation

Epic files FTC complaint over Samsung smart TV voice recording accusing the company of breachi laws. Photograph: Kārlis Dambrāns/flickr

TechHive FROM IDG

all sections
NEWS • REVIEWS • SMART HOME • ENTERTAINMENT • DEALS • RESOURCES
Music Streaming Media
Start Searching

NEWS

Vizio smart TVs track what you watch, sell the behavior data

Behavior tracking is enabled by default and lets advertisers follow you beyond the television. But it's not much worse than the status quo.

By Jared Newman
TechHive | NOV 10, 2015 7:48 AM PST

'Smart' home devices used as weapons in website attack

October 2016 | Technology

Net-connected cameras are helping attackers in large-scale attacks

Hackers used internet-connected home devices, such as CCTV cameras and printers, to attack popular websites on Friday, security analysts say.

BBC Sign in News Sport Weather Shop Earth Travel Mor
NEWS
Home Video World UK Business Tech Science Magazine Entertainment & Arts

Technology

Warning over smart meters privacy risk

12 June 2012 | Technology

An EU data watchdog has warned of the "considerable risks" to privacy posed by new energy smart meters.

The European Data Protection Supervisor said safeguards were needed over how firms used the "massive collection" of consumers' data uploaded by meters.

Smart meters are able to upload data about consumers' energy use to third parties

ED MARKEY
United States Senator *for* Massachusetts

Home / News / Press Releases / Press Release

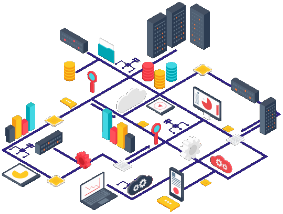
Markey Report Reveals Automobile Security and Privacy Vulnerabilities

Monday, February 9, 2015

Wireless technologies leave vehicles exposed to hackers; Information collected on driver locations, habits

WASHINGTON (February 9, 2014) – New standards are needed to plug security and privacy gaps in our cars and trucks, according to a report released today by Senator Edward J. Markey (D-Mass.). The report, called *Tracking & Hacking: Security & Privacy Gaps Put American Drivers at Risk* and first reported on by CBS News' 60 Minutes, reveals how sixteen major automobile manufacturers responded to questions from Senator Markey in 2014 about how vehicles may be vulnerable to hackers, and how driver information is collected and protected.

Security . . . a complex problem



Protection of infrastructure

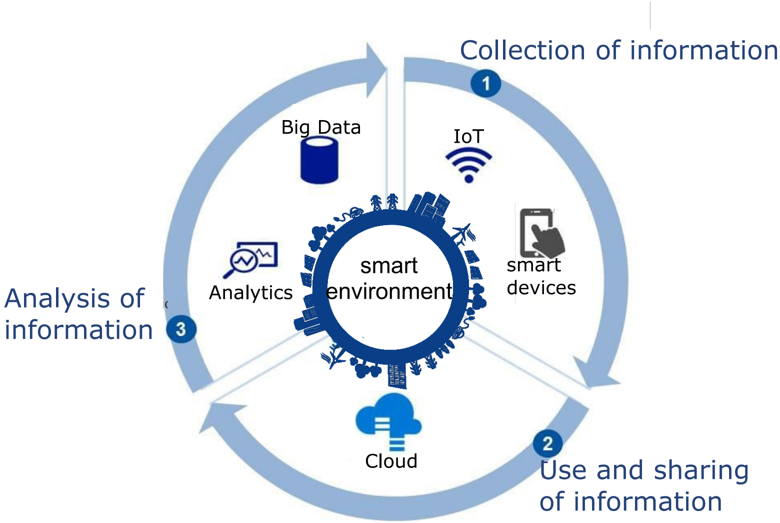
Protection of communication

Protection against malware and attacks

Protection of devices

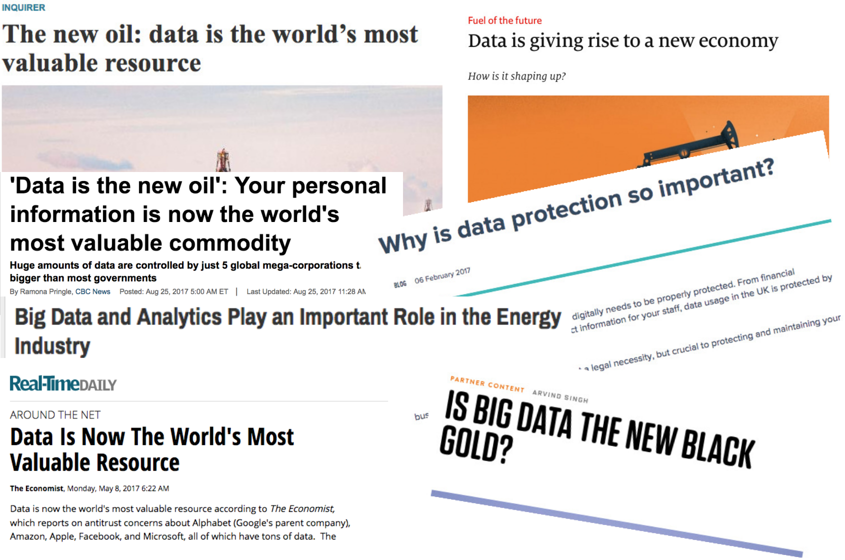
Protection of data

The role of data in a smart environment



⟹ better governance and intelligent systems

The most valuable resource - Data



INQUIRER

The new oil: data is the world's most valuable resource

'Data is the new oil': Your personal information is now the world's most valuable commodity

Huge amounts of data are controlled by just 5 global mega-corporations t
bigger than most governments

By Ramona Pringle, CBC News Posted: Aug 25, 2017 5:00 AM ET | Last Updated: Aug 25, 2017 11:28 AN

Fuel of the future

Data is giving rise to a new economy

How is it shaping up?

Why is data protection so important?

BLOG 06 February 2017

digitally needs to be properly protected. From financial
ct information for your staff, data usage in the UK is protected by

a legal necessity, but crucial to protecting and maintaining your

bus

Big Data and Analytics Play an Important Role in the Energy Industry

Real-Time DAILY

AROUND THE NET

Data Is Now The World's Most Valuable Resource

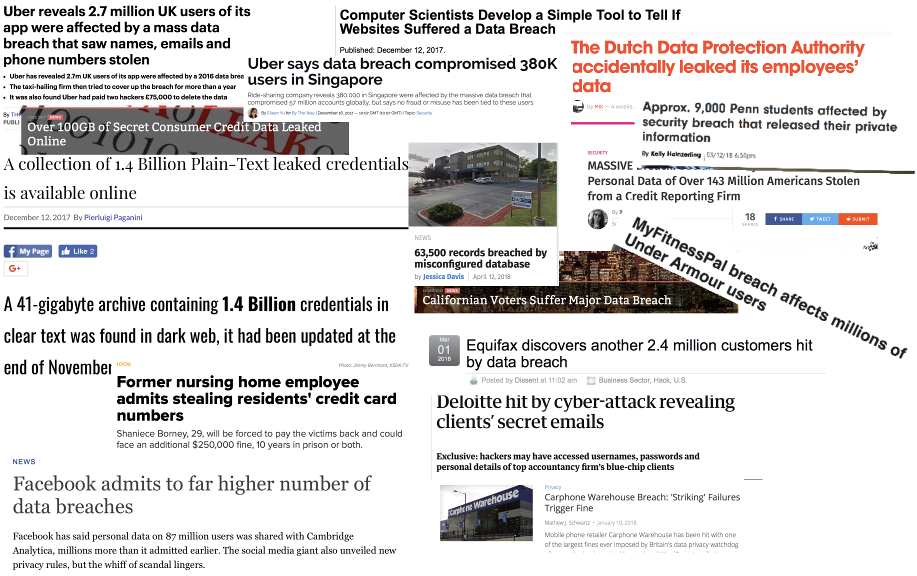
The Economist, Monday, May 8, 2017 6:22 AM

Data is now the world's most valuable resource according to *The Economist,* which reports on antitrust concerns about Alphabet (Google's parent company), Amazon, Apple, Facebook, and Microsoft, all of which have tons of data. The

PARTNER CONTENT ARVIND SINGH

IS BIG DATA THE NEW BLACK GOLD?

Impact on data protection and privacy



Uber reveals 2.7 million UK users of its app were affected by a mass data breach that saw names, emails and phone numbers stolen

- Uber has revealed 2.7m UK users of its app were affected by a 2016 data brea
- The taxi-hailing firm then tried to cover up the breach for more than a year
- It was also found Uber had paid two hackers £75,000 to delete the data

By TH
PUBLI

Computer Scientists Develop a Simple Tool to Tell If Websites Suffered a Data Breach

Published: December 12, 2017.

Uber says data breach compromised 380K users in Singapore

Ride-sharing company reveals 380,000 in Singapore were affected by the massive data breach that compromised 57 million accounts globally, but says no fraud or misuse has been tied to these users.

By Eileen Yu for By The Way | December 16, 2017 -- 10:07 GMT (10:07 GMT) | Topic: Security

The Dutch Data Protection Authority accidentally leaked its employees' data

by MIX — 4 weeks

1 DEC 2017 NEWS

Over 100GB of Secret Consumer Credit Data Leaked Online

A collection of 1.4 Billion Plain-Text leaked credentials is available online

December 12, 2017 By Pierluigi Paganini

My Page | Like 2 | G+

A 41-gigabyte archive containing **1.4 Billion** credentials in clear text was found in dark web, it had been updated at the end of November

Approx. 9,000 Penn students affected by security breach that released their private information

By Kelly Heinzerling | 03/12/18 6:50pm

SECURITY

MASSIVE

Personal Data of Over 143 Million Americans Stolen from a Credit Reporting Firm

By R

18 SHARES | SHARE | TWEET | SUBMIT

NEWS

63,500 records breached by misconfigured database

by Jessica Davis | April 12, 2018

MyFitnessPal breach affects millions of Under Armour users

19 DEC 2017 NEWS

Californian Voters Suffer Major Data Breach

Mar 01 2018

Equifax discovers another 2.4 million customers hit by data breach

Posted by Dissent at 11:02 am | Business Sector, Hack, U.S.

LOCAL

Photo: Jimmy Bernhard, KSDK-TV

Former nursing home employee admits stealing residents' credit card numbers

Shaniece Borney, 29, will be forced to pay the victims back and could face an additional $250,000 fine, 10 years in prison or both.

Deloitte hit by cyber-attack revealing clients' secret emails

Exclusive: hackers may have accessed usernames, passwords and personal details of top accountancy firm's blue-chip clients

NEWS

Facebook admits to far higher number of data breaches

Facebook has said personal data on 87 million users was shared with Cambridge Analytica, millions more than it admitted earlier. The social media giant also unveiled new privacy rules, but the whiff of scandal lingers.

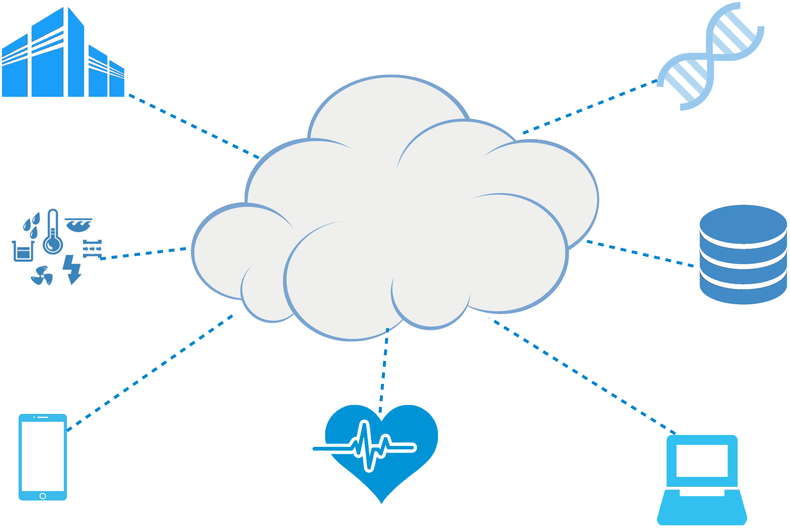
Privacy

Carphone Warehouse Breach: 'Striking' Failures Trigger Fine

Mathew J. Schwartz • January 10, 2018

Mobile phone retailer Carphone Warehouse has been hit with one of the largest fines ever imposed by Britain's data privacy watchdog

Huge amount of data stored at external providers

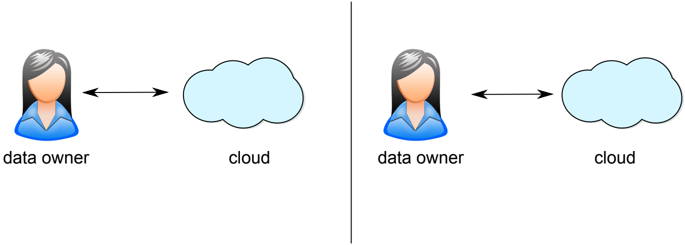


Cloud computing

- The Cloud allows users and organizations to rely on external providers for storing, processing, and accessing their data
  - + high configurability and economy of scale
  - + data and services are always available
  - + scalable infrastructure for applications
- Users lose control over their own data
  - – new security and privacy problems
- Need solutions to protect data and to securely process them in the cloud

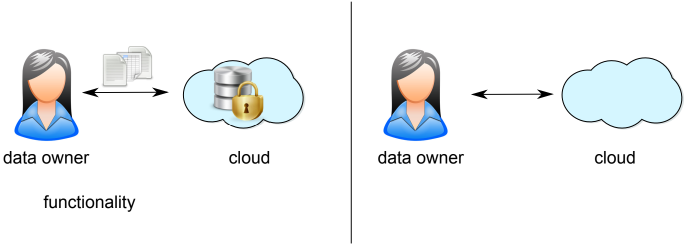
Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



Cloud computing: Today

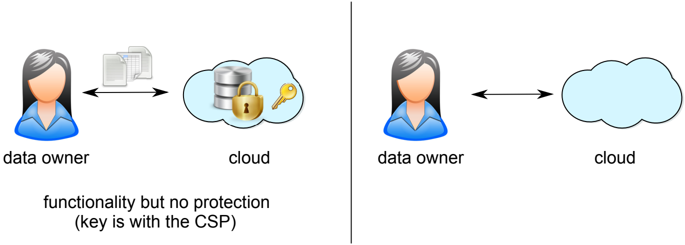
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



- functionality

Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



data owner

cloud

functionality but no protection
(key is with the CSP)

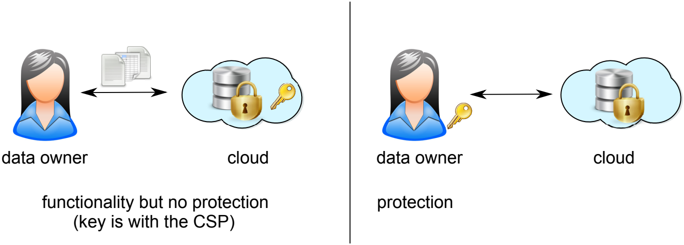
data owner

cloud

- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)

Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



data owner

cloud

functionality but no protection
(key is with the CSP)

data owner

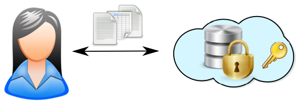
cloud

protection

- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)
- protection

Cloud computing: Today

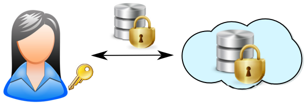
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



data owner

cloud

functionality but no protection
(key is with the CSP)



data owner

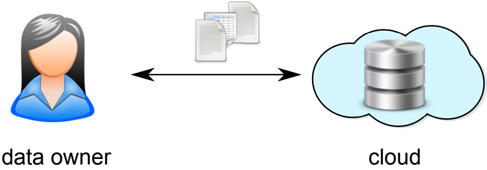
cloud

protection but limited functionality
(you cannot access data as you like)

- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)
- protection but limited functionality since the CSP cannot access data (e.g., Boxcryptor, SpiderOak)

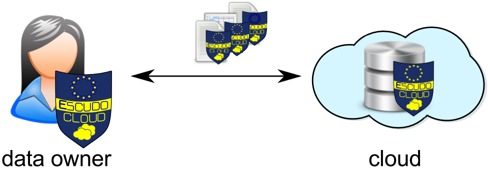
Cloud computing: New vision

Solutions that provide protection guarantees giving the data owners both: full control over their data and cloud functionality over them



Cloud computing: New vision

Solutions that provide protection guarantees giving the data owners both: full control over their data and cloud functionality over them

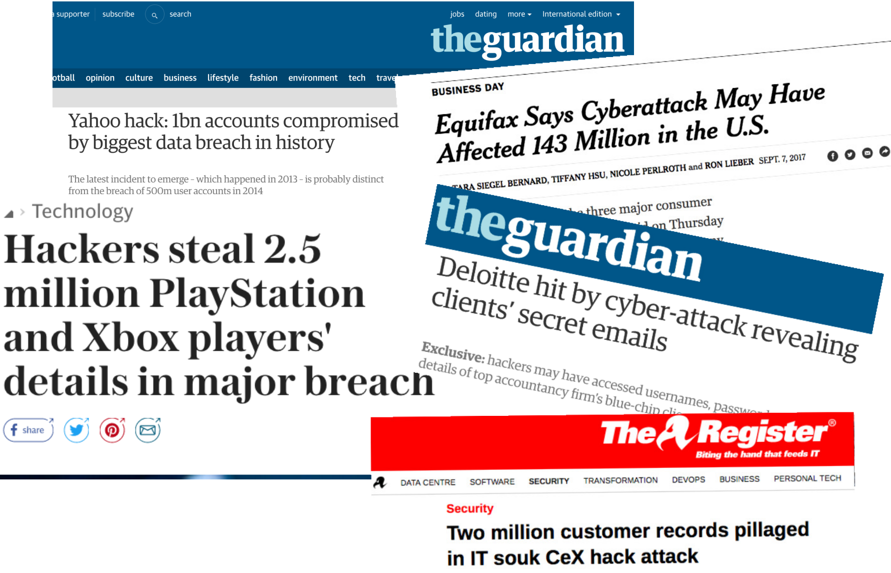


- client-side trust boundary: only the behavior of the client should be considered trusted

  ⟹ techniques and implementations supporting direct processing of encrypted data in the cloud

https://www.marsalproject.eu https://glaciation-project.eu SERICS https://serics.eu

Data protection – Base level



supporter | subscribe | search | jobs | dating | more | International edition

theguardian

otball | opinion | culture | business | lifestyle | fashion | environment | tech | travel

Yahoo hack: 1bn accounts compromised by biggest data breach in history

The latest incident to emerge - which happened in 2013 - is probably distinct from the breach of 500m user accounts in 2014

› Technology

Hackers steal 2.5 million PlayStation and Xbox players' details in major breach

share

BUSINESS DAY

Equifax Says Cyberattack May Have Affected 143 Million in the U.S.

TARA SIEGEL BERNARD, TIFFANY HSU, NICOLE PERLROTH and RON LIEBER SEPT. 7, 2017

three major consumer … on Thursday

theguardian

Deloitte hit by cyber-attack revealing clients' secret emails

Exclusive: hackers may have accessed usernames, passwo… details of top accountancy firm's blue-chip cli…

The Register®

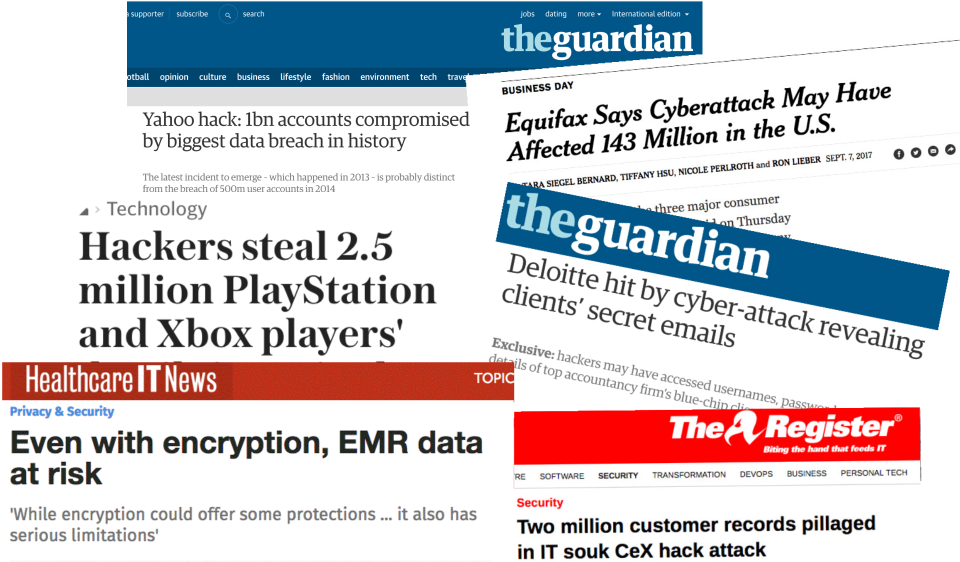
Biting the hand that feeds IT

DATA CENTRE | SOFTWARE | SECURITY | TRANSFORMATION | DEVOPS | BUSINESS | PERSONAL TECH

Security

Two million customer records pillaged in IT souk CeX hack attack

Data protection – Base level



a supporter subscribe search jobs dating more International edition

theguardian

otball opinion culture business lifestyle fashion environment tech trave

Yahoo hack: 1bn accounts compromised by biggest data breach in history

The latest incident to emerge - which happened in 2013 - is probably distinct from the breach of 500m user accounts in 2014

Technology

Hackers steal 2.5 million PlayStation and Xbox players'

BUSINESS DAY

Equifax Says Cyberattack May Have Affected 143 Million in the U.S.

TARA SIEGEL BERNARD, TIFFANY HSU, NICOLE PERLROTH and RON LIEBER SEPT. 7, 2017

three major consumer

on Thursday

theguardian

Deloitte hit by cyber-attack revealing clients' secret emails

Exclusive: hackers may have accessed usernames, passwo

details of top accountancy firm's blue-chip cli

Healthcare IT News TOPIC

Privacy & Security

Even with encryption, EMR data at risk

'While encryption could offer some protections ... it also has serious limitations'

The Register®

Biting the hand that feeds IT

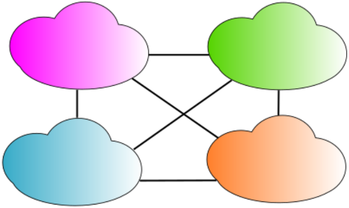
RE SOFTWARE SECURITY TRANSFORMATION DEVOPS BUSINESS PERSONAL TECH

Security

Two million customer records pillaged in IT souk CeX hack attack

Data protection – Regulation

Access and usage control

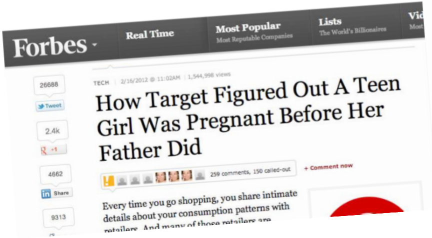


Selective sharing

Governance and regulation

Data protection – Confidentiality (1)

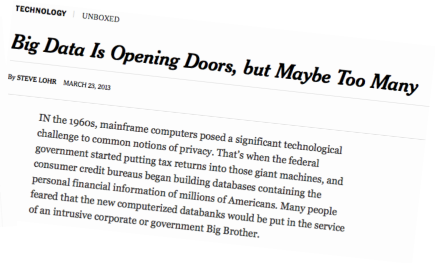
- Minimize release/exposition
  - ○ correlation among different data sources
  - ○ indirect exposure of sensitive information
  - ○ de-identification $\neq$ anonymization



TECH | 2/16/2012 @ 11:02AM | 1,544,998 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are

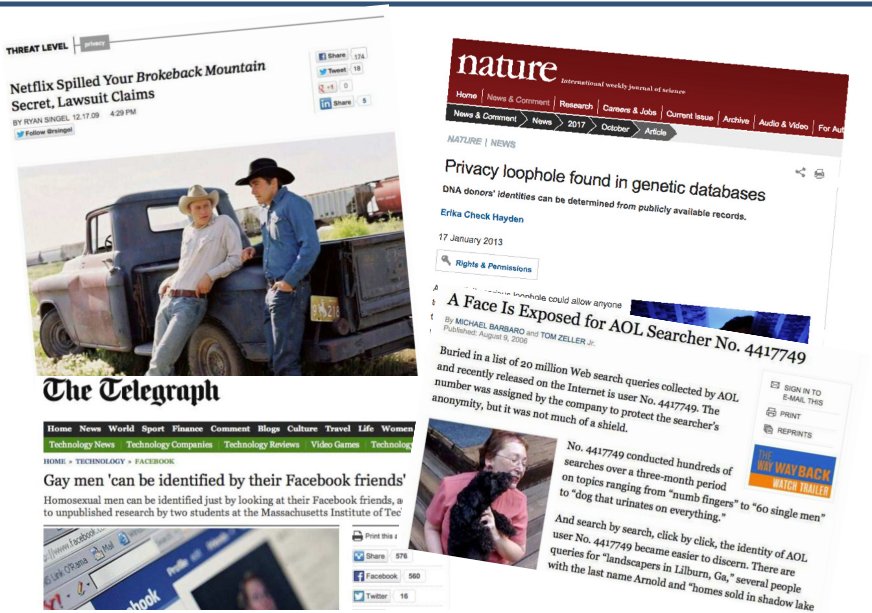


Big Data Is Opening Doors, but Maybe Too Many

By STEVE LOHR MARCH 23, 2013

IN the 1960s, mainframe computers posed a significant technological challenge to common notions of privacy. That's when the federal government started putting tax returns into those giant machines, and consumer credit bureaus began building databases containing the personal financial information of millions of Americans. Many people feared that the new computerized databanks would be put in the service of an intrusive corporate or government Big Brother.

Data protection – Confidentiality (2)


THREAT LEVEL
privacy
Share 174
Tweet 18
+1 0
Share 5
Follow @rsingel
9 218
nature
International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Aut
News & Comment > News > 2017 > October > Article
NATURE | NEWS
Rights & Permissions
SIGN IN TO E-MAIL THIS
PRINT
REPRINTS
THE WAY WAY BACK
WATCH TRAILER
The Telegraph
Home News World Sport Finance Comment Blogs Culture Travel Life Women
Technology News | Technology Companies | Technology Reviews | Video Games | Technolog
HOME » TECHNOLOGY » FACEBOOK
Print this a
Share 576
Facebook 560
Twitter 16
www.facebook.co
Mail
Link O'Rama
Profile
ebook

Netflix Spilled Your *Brokeback Mountain* Secret, Lawsuit Claims

BY RYAN SINGEL 12.17.09 4:29 PM

Privacy loophole found in genetic databases

DNA donors' identities can be determined from publicly available records.

Erika Check Hayden

17 January 2013

A ... loophole could allow anyone

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake

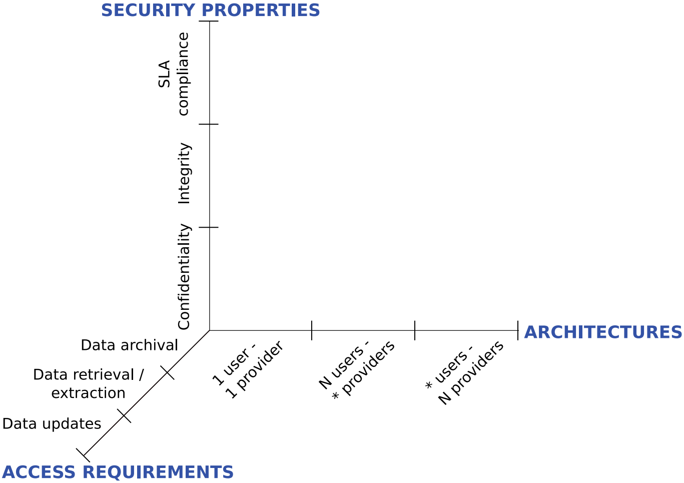
Gay men 'can be identified by their Facebook friends'

Homosexual men can be identified just by looking at their Facebook friends, a
to unpublished research by two students at the Massachusetts Institute of Tec

Characterization of Data Protection Challenges in Cloud Scenarios

Scientific and technical challenges

Three dimensions characterize the problems and challenges



Security properties

Confidentiality

- data externally stored
- users identities
- actions that users perform on the data

Integrity

- data externally stored
- computation and query results

SLA compliance

- assurance and certification

Access requirements

Data archival

- upload/download
- protection of data in storage

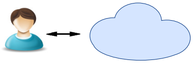
Data retrieval/extraction

- support for fine-grained data retrieval and queries
- protection of computations and query results

Data update

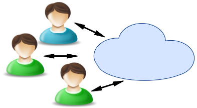
- support for access retrieval and enforcement of updates
- protection of the actions and of their effects on the data

Architectures



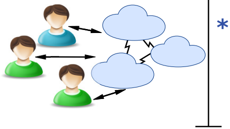
1 user - 1 provider

- protection of data at rest
- fine-grained retrieval
- query privacy/integrity



n users - * providers

- authorizations and access control
- multiple writers



*** users - n providers**

- controlled data sharing and computation

Combinations of the dimensions

- Every combination of the different instances of the dimensions identifies new problems and challenges

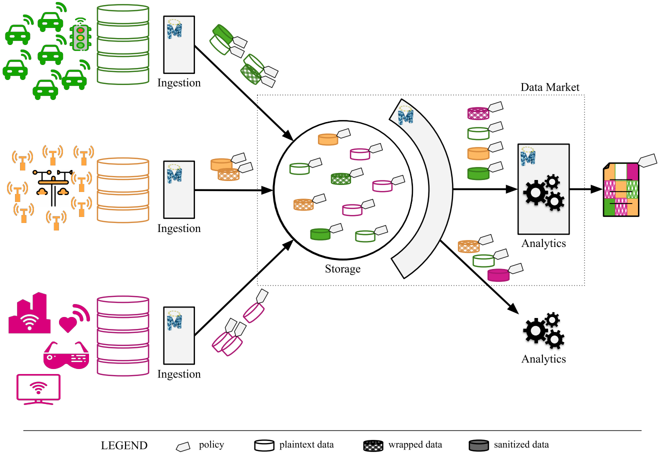
- The security properties to be guaranteed can depend on the access requirements and on the trust assumption on the providers involved in storage and/or processing of data

- Providers can be:
  - curious
  - lazy
  - malicious

Digital Data Market

Goal and vision

Enable data sharing and collaborative computations in multi-provider / multi-owner scenarios, while ensuring proper protection of sensitive or company-confidential information



Dimensions of the problems and challenges

- Requirements capturing and representation

  policies regulating access, sharing, usage and processing

- Enforcing technologies

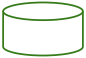
  data wrapping / sanitization

- Enforcement phase

  ingestion / storage / analytics

Requirements capturing and representation

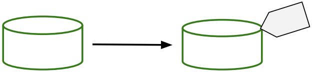
Data owners need to have a way to express their requirements and having them enforced



Requirements capturing and representation

Data owners need to have a way to express their requirements and having them enforced

- Policies regulate access, sharing, usage and processing of data



Enforcing technologies

Techniques and mechanisms for enforcing data protection

Enforcing technologies

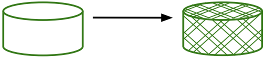
Techniques and mechanisms for enforcing data protection

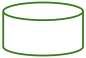
- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

Enforcing technologies

Techniques and mechanisms for enforcing data protection

- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

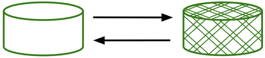


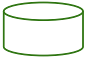


Enforcing technologies

Techniques and mechanisms for enforcing data protection

- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

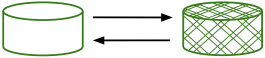




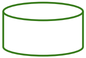
Enforcing technologies

Techniques and mechanisms for enforcing data protection

- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality



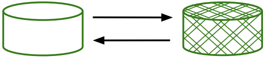
- Sanitization: provide protection by returning an obfuscated (e.g., not precise) version of the data



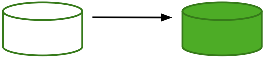
Enforcing technologies

Techniques and mechanisms for enforcing data protection

- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality



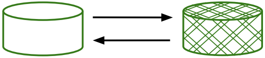
- Sanitization: provide protection by returning an obfuscated (e.g., not precise) version of the data



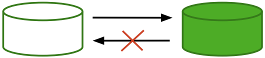
Enforcing technologies

Techniques and mechanisms for enforcing data protection

- Wrapping: provide protection by (partially or completely) disabling visibility of data while preserving some functionality

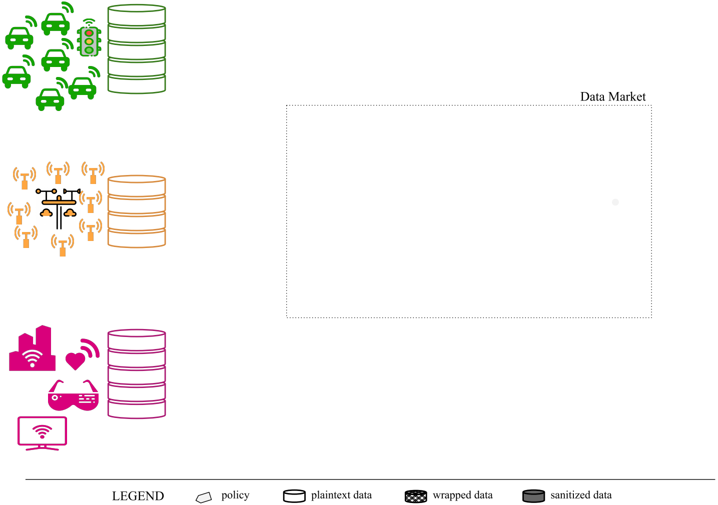


- Sanitization: provide protection by returning an obfuscated (e.g., not precise) version of the data



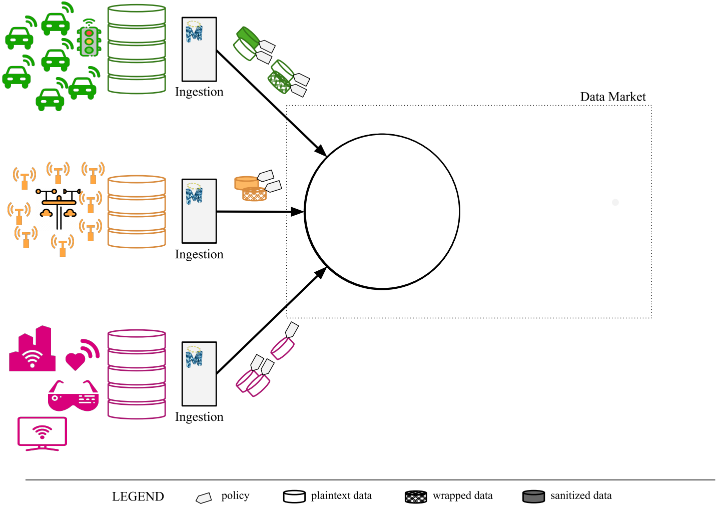
Enforcement phase

- Ingestion / Storage / Analytics



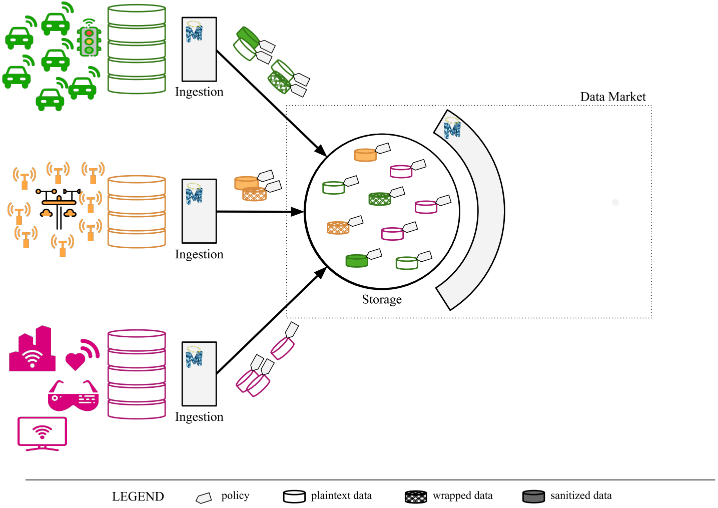
Enforcement phase

- Ingestion / Storage / Analytics



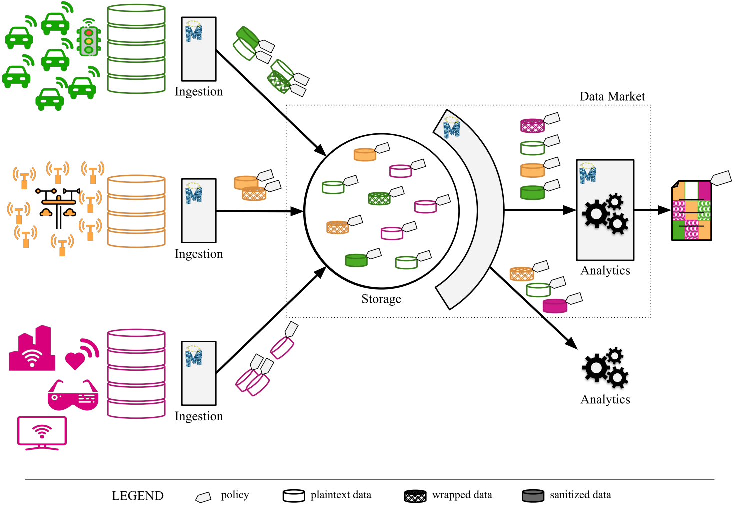
Enforcement phase

- Ingestion / Storage / Analytics



Enforcement phase

- Ingestion / Storage / Analytics



Some open issues

Controlled collaborative query execution

Fine-grained access over encrypted data

Distributed resource allocation and computation

Secure energy-aware data management

Providers/plans selection

Access confidentiality

User privacy

Security metrics

Computation integrity

Protection of data at rest

Query privacy

Data publication and utility

Green IT and cybersecurity

Policy definition and modeling

Some open issues

Controlled collaborative query execution

Fine-grained access over encrypted data

Distributed resource allocation and computation

Secure energy-aware data management

Providers/plans selection

Access confidentiality

User privacy

Security metrics

Computation integrity

Protection of data at rest

Query privacy

Data publication and utility

Green IT and cybersecurity

Policy definition and modeling

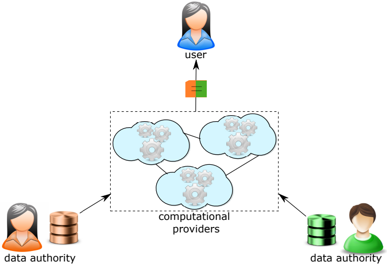
Controlled Collaborative Query Execution

S. De Capitani di Vimercati, S. Foresti, S. Jajodia, G. Livraga, S. Paraboschi, P. Samarati, "An Authorization Model for Query Execution in the Cloud," in *The VLDB Journal*, vol. 31, n. 3, May 2022, pp. 555-579.

Data markets

- Represent a promising solution for combining data from different sources
- Store data of different owners that could be sensitive, proprietary, or subject to access restrictions
- Participate and partially delegate query evaluation to third parties

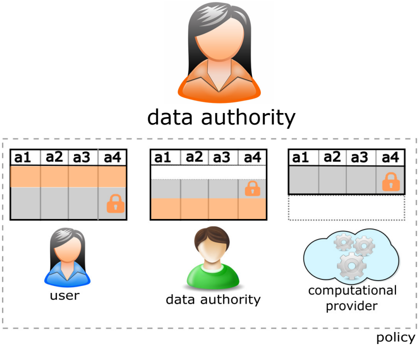
⟹ Need solutions for supporting controlled collaborative query execution



Challenges: Policies

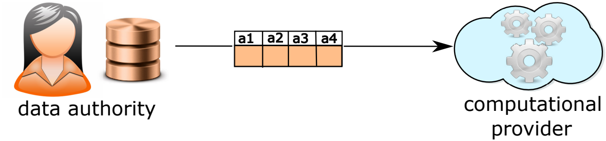
- Data could be sensitive, proprietary, or subject to access restrictions

- Need to define policies to regulate data visibility



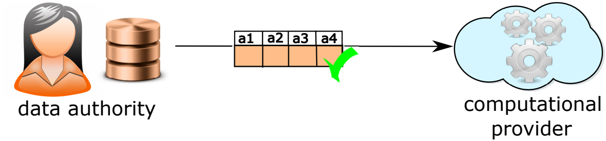
Challenges: Information flows

- Need to ensure no information is directly or indirectly leaked in the execution process



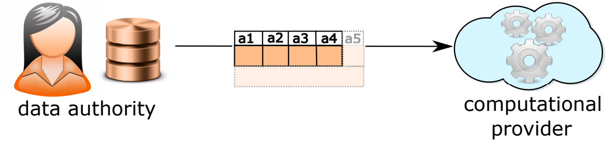
Challenges: Information flows

- Need to ensure no information is directly or indirectly leaked in the execution process



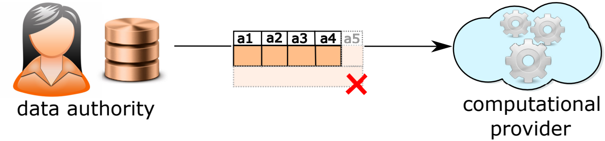
Challenges: Information flows

- Need to ensure no information is directly or indirectly leaked in the execution process



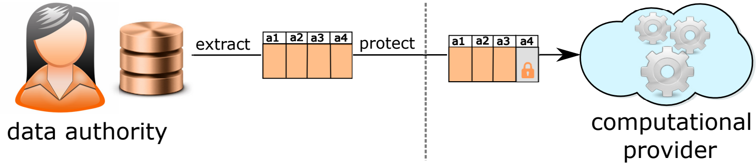
Challenges: Information flows

- Need to ensure no information is directly or indirectly leaked in the execution process



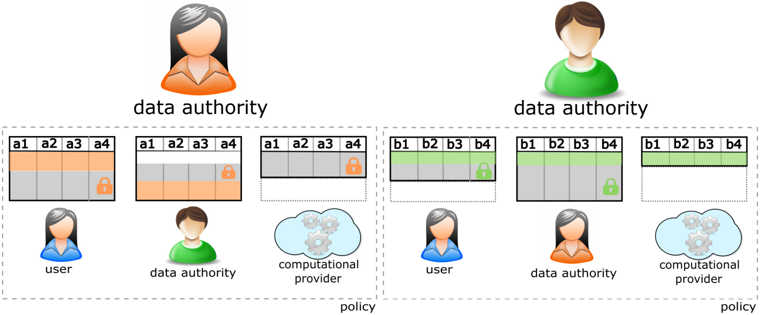
Challenges: Policy enforcement

- Need solutions for dynamically protect sensitive/confidential information as needed



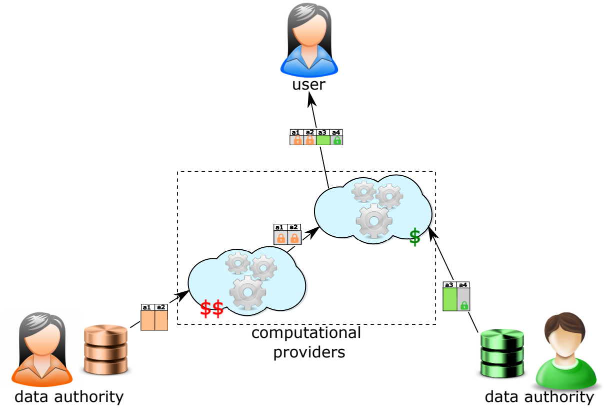
Challenges: Independency

- Authorities/data owners need to independently specify the policies regulating access to their own data



Challenges: Preference factors

- Need to support the selective involvement of external providers when convenient (e.g., economically) while preserving data confidentiality



Some existing approaches

- Sovereign joins
- Access patterns
- View-based access control
- Authorizations with join paths for enabling distributed query evaluation
- …
- Controlled data sharing for collaborative queries in the cloud

Controlled data sharing for collaborative queries

- Simple yet flexible authorization model
- Plaintext/encrypted visibility over attributes
- Authorities make data available, while maintaining control
- Users can involve external providers for query evaluation while preserving data confidentiality

Authorization model

- Authorities specify authorizations on their relations granting access to attributes in two forms: plaintext and encrypted

Authorization model

- Authorities specify authorizations on their relations granting access to attributes in two forms: plaintext and encrypted

	Relation					
Subject	Hosp@ℍ				Ins@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Authorization model

- Authorities specify authorizations on their relations granting access to attributes in two forms: plaintext and encrypted
- Given a query plan, a set of cloud providers, and a set of authorizations, compute an authorized assignment

	Relation	
Subject	HOSP@ℍ	INS@𝕀
ℍ	S B D T	C P
𝕀	S B D T	C P
𝕌	S D T	C P
𝕏	S D T	C P
𝕐	S B D T	C P
ℤ	S D T	C P

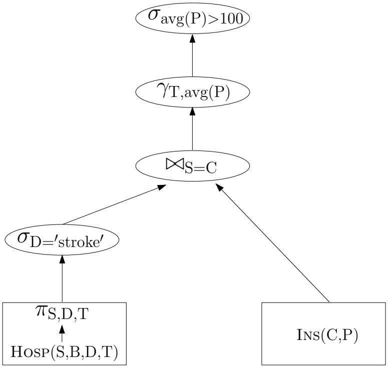
HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Authorization model

- Authorities specify authorizations on their relations granting access to attributes in two forms: plaintext and encrypted
- Given a query plan, a set of cloud providers, and a set of authorizations, compute an authorized assignment

		Relation					
		HOSP@ℍ				INS@𝕀	
Subject	ℍ	S	B	D	T	C	P
	𝕀	S	B	D	T	C	P
	𝕌	S		D	T	C	P
	𝕏	S		D	T	C	P
	𝕐	S	B	D	T	C	P
	ℤ	S		D	T	C	P

```
SELECT      T, avg(P)
FROM        HOSP JOIN INS ON S=C
WHERE       D='stroke'
GROUP BY    T
HAVING      avg(P)>100
```



HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Relation profile

- Captures information content of a relation R and includes

Relation profile

- Captures information content of a relation R and includes

  v : visible attributes: plaintext or encrypted in R's schema

Relation profile

- Captures information content of a relation R and includes

  v : visible attributes: plaintext or encrypted in R's schema

  i : implicit attributes: conveyed, plaintext or encrypted, by R

  – selection: SELECT S FROM HOSP WHERE D='stroke'
    leaks the value of D, even if D does not belong to the schema

  – grouping: SELECT COUNT(*) FROM HOSP JOIN INS ON S=C GROUP BY T
    leaks information on tuples with the same value for T, even if T does not belong to the schema

Relation profile

- Captures information content of a relation R and includes

  v : visible attributes: plaintext or encrypted in R's schema

  i : implicit attributes: conveyed, plaintext or encrypted, by R

  - selection: SELECT S FROM HOSP WHERE D='stroke'
    leaks the value of D, even if D does not belong to the schema
  - grouping: SELECT COUNT(*) FROM HOSP JOIN INS ON S=C GROUP BY T
    leaks information on tuples with the same value for T, even if T does not belong to the schema

  $\simeq$: equivalence relationship: among attributes connected in R's computation

  - comparing attributes: SELECT S FROM HOSP JOIN INS ON S=C
    leaks the values of C, even if C does not belong to the schema

Relation profile

- Captures information content of a relation R and includes

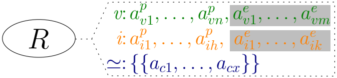
  v : visible attributes: plaintext or encrypted in R's schema

  i : implicit attributes: conveyed, plaintext or encrypted, by R

  - selection: SELECT S FROM HOSP WHERE D=‘stroke’
    leaks the value of D, even if D does not belong to the schema
  - grouping: SELECT COUNT(*) FROM HOSP JOIN INS ON S=C GROUP BY T
    leaks information on tuples with the same value for T, even if T does not belong to the schema

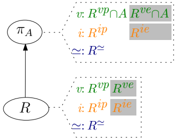
  $\simeq$: equivalence relationship: among attributes connected in R's computation

  - comparing attributes: SELECT S FROM HOSP JOIN INS ON S=C
    leaks the values of C, even if C does not belong to the schema


R
v: $a^p_{v1}, \ldots, a^p_{vn}, a^e_{v1}, \ldots, a^e_{vm}$
i: $a^p_{i1}, \ldots, a^p_{ih}, a^e_{i1}, \ldots, a^e_{ik}$
$\simeq$: $\{\{a_{c1}, \ldots, a_{cx}\}\}$

Profiles resulting from operations

Projection


π_A

v: $R^{vp}\cap A$ $R^{ve}\cap A$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

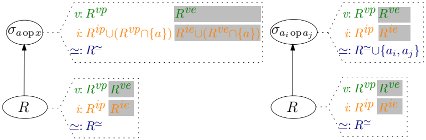
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

Selection


$\sigma_{a\,\text{op}\,x}$

v: R^{vp} R^{ve}

i: $R^{ip}\cup(R^{vp}\cap\{a\})$ $R^{ie}\cup(R^{ve}\cap\{a\})$

$\simeq$: $R^{\simeq}$

R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

$\sigma_{a_i\,\text{op}\,a_j}$

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}\cup\{a_i, a_j\}$

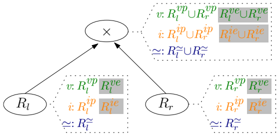
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

Cartesian Product


$\times$

v: $R_l^{vp}\cup R_r^{vp}$ $R_l^{ve}\cup R_r^{ve}$

i: $R_l^{ip}\cup R_r^{ip}$ $R_l^{ie}\cup R_r^{ie}$

$\simeq$: $R_l^{\simeq}\cup R_r^{\simeq}$

R_l

v: R_l^{vp} R_l^{ve}

i: R_l^{ip} R_l^{ie}

$\simeq$: $R_l^{\simeq}$

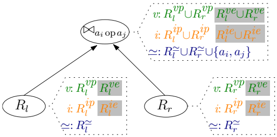
R_r

v: R_r^{vp} R_r^{ve}

i: R_r^{ip} R_r^{ie}

$\simeq$: $R_r^{\simeq}$

Join


$\bowtie_{a_i\,\text{op}\,a_j}$

v: $R_l^{vp}\cup R_r^{vp}$ $R_l^{ve}\cup R_r^{ve}$

i: $R_l^{ip}\cup R_r^{ip}$ $R_l^{ie}\cup R_r^{ie}$

$\simeq$: $R_l^{\simeq}\cup R_r^{\simeq}\cup\{a_i, a_j\}$

R_l

v: R_l^{vp} R_l^{ve}

i: R_l^{ip} R_l^{ie}

$\simeq$: $R_l^{\simeq}$

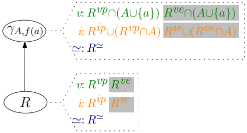
R_r

v: R_r^{vp} R_r^{ve}

i: R_r^{ip} R_r^{ie}

$\simeq$: $R_r^{\simeq}$

Group by


$\gamma_{A,f(a)}$

v: $R^{vp}\cap(A\cup\{a\})$ $R^{ve}\cap(A\cup\{a\})$

i: $R^{ip}\cup(R^{vp}\cap A)$ $R^{ie}\cup(R^{ve}\cap A)$

$\simeq$: $R^{\simeq}$

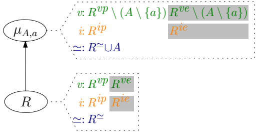
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

User defined functions


$\mu_{A,a}$

v: $R^{vp}\setminus(A\setminus\{a\})$ $R^{ve}\setminus(A\setminus\{a\})$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}\cup A$

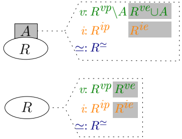
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

Encryption


A

R

v: $R^{vp}\setminus A$ $R^{ve}\cup A$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

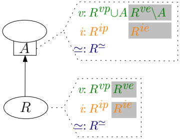
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

Decryption


A

v: $R^{vp}\cup A$ $R^{ve}\setminus A$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

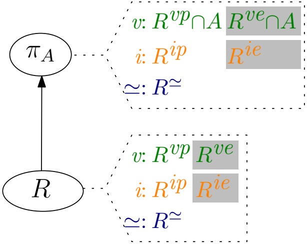
R

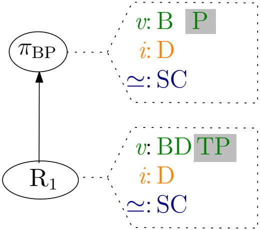
v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

Projection



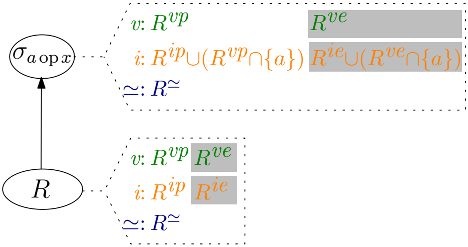


SELECT A
FROM R

SELECT B, P
FROM R_1

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Selection – 1



$\sigma_{a\,\mathrm{op}\,x}$

v: R^{vp} R^{ve}

i: $R^{ip} \cup (R^{vp} \cap \{a\})$ $R^{ie} \cup (R^{ve} \cap \{a\})$

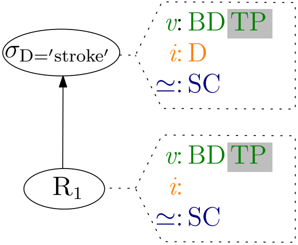
$\simeq$: $R^{\simeq}$

R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$



$\sigma_{\mathrm{D}='\mathrm{stroke}'}$

v: BD TP

i: D

$\simeq$: SC

R_1

v: BD TP

i:

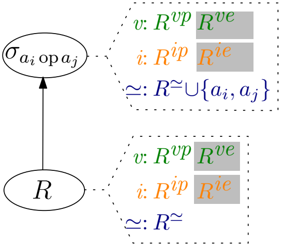
$\simeq$: SC

SELECT $*$
FROM R
WHERE a op x

SELECT $*$
FROM R_1
WHERE D='stroke'

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Selection – 2



$\sigma_{a_i \,\mathrm{op}\, a_j}$

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

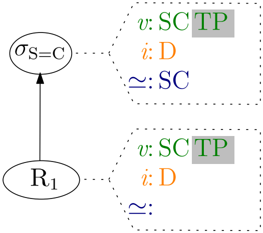
$\simeq$: $R^{\simeq} \cup \{a_i, a_j\}$

R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$



$\sigma_{\mathrm{S=C}}$

v: SC TP

i: D

$\simeq$: SC

R_1

v: SC TP

i: D

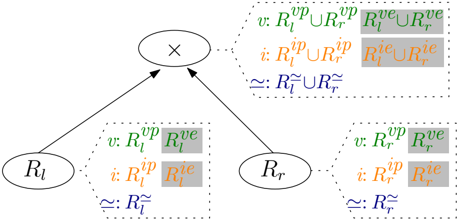
$\simeq$:

SELECT $*$
FROM R
WHERE a_i op a_j

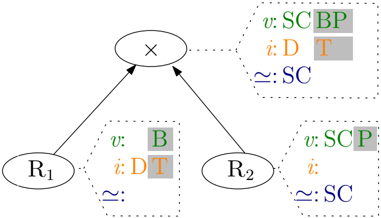
SELECT $*$
FROM R_1
WHERE S=C

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Cartesian product



$\times$
v: $R_l^{vp} \cup R_r^{vp}$ $R_l^{ve} \cup R_r^{ve}$
i: $R_l^{ip} \cup R_r^{ip}$ $R_l^{ie} \cup R_r^{ie}$
$\simeq$: $R_l^{\simeq} \cup R_r^{\simeq}$
R_l
v: R_l^{vp} R_l^{ve}
i: R_l^{ip} R_l^{ie}
$\simeq$: $R_l^{\simeq}$
R_r
v: R_r^{vp} R_r^{ve}
i: R_r^{ip} R_r^{ie}
$\simeq$: $R_r^{\simeq}$

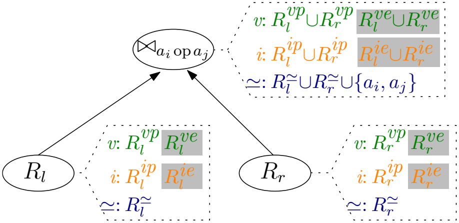


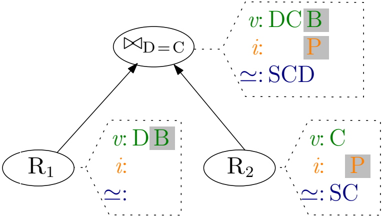
SELECT $*$
FROM $R_l \times R_r$

SELECT $*$
FROM $\mathrm{R}_1 \times \mathrm{R}_2$

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment)   INS(**C**ustomer, **P**remium)

Join


$\bowtie_{a_i\,\mathrm{op}\,a_j}$
v: $R_l^{vp} \cup R_r^{vp}$ $R_l^{ve} \cup R_r^{ve}$
i: $R_l^{ip} \cup R_r^{ip}$ $R_l^{ie} \cup R_r^{ie}$
$\simeq$: $R_l^{\simeq} \cup R_r^{\simeq} \cup \{a_i, a_j\}$
R_l
v: R_l^{vp} R_l^{ve}
i: R_l^{ip} R_l^{ie}
$\simeq$: $R_l^{\simeq}$
R_r
v: R_r^{vp} R_r^{ve}
i: R_r^{ip} R_r^{ie}
$\simeq$: $R_r^{\simeq}$

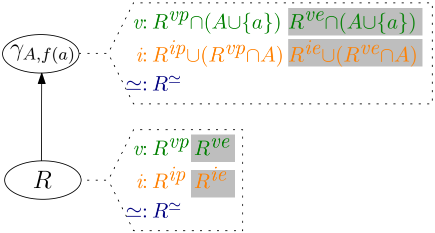


SELECT $*$
FROM R_l JOIN R_r ON a_i op a_j

SELECT $*$
FROM R_1 JOIN R_2
ON S=C

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Group by



$\gamma_{A,f(a)}$

v: $R^{vp}\cap(A\cup\{a\})$ $R^{ve}\cap(A\cup\{a\})$

i: $R^{ip}\cup(R^{vp}\cap A)$ $R^{ie}\cup(R^{ve}\cap A)$

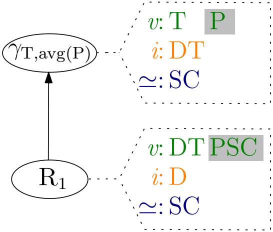
$\simeq$: $R^{\simeq}$

R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$



$\gamma_{\mathrm{T,avg(P)}}$

v: T P

i: DT

$\simeq$: SC

R_1

v: DT PSC

i: D

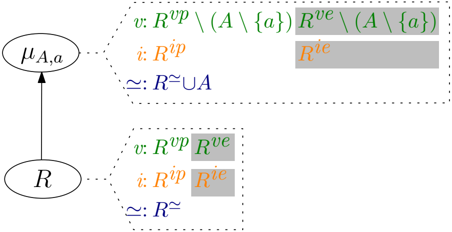
$\simeq$: SC

SELECT $A,f(a)$
FROM R
GROUP BY A

SELECT T,AVG(P)
FROM R_1
GROUP BY T

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

User defined functions



$\mu_{A,a}$

v: $R^{vp} \setminus (A \setminus \{a\})$ $R^{ve} \setminus (A \setminus \{a\})$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq} \cup A$

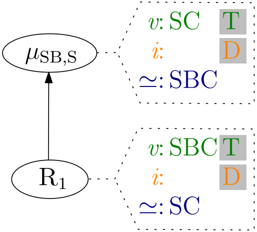
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

a AS UDF(A)



$\mu_{\text{SB,S}}$

v: SC T

i: D

$\simeq$: SBC

R_1

v: SBC T

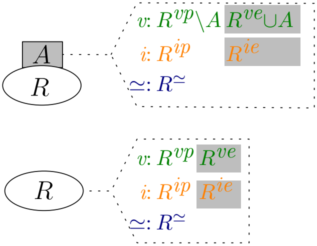
i: D

$\simeq$: SC

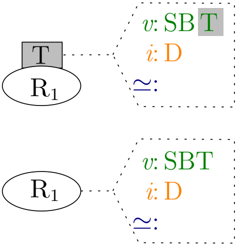
S AS UDF(S,B)

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Encryption



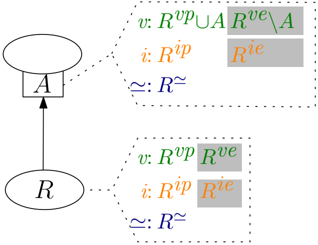
ENCRYPT$(R.A)$



ENCRYPT$(\mathrm{R}_1.\mathrm{T})$

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Decryption



v: $R^{vp} \cup A$ $R^{ve} \setminus A$

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

A

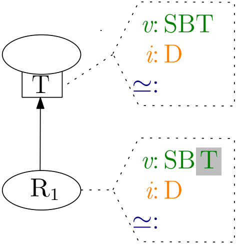
R

v: R^{vp} R^{ve}

i: R^{ip} R^{ie}

$\simeq$: $R^{\simeq}$

DECRYPT$(R.A)$



v: SBT

i: D

$\simeq$:

T

R_1

v: SBT

i: D

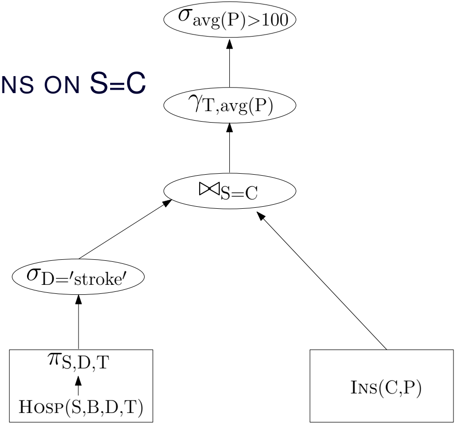
$\simeq$:

DECRYPT$(\mathrm{R}_1.\mathrm{T})$

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Plan with profiles

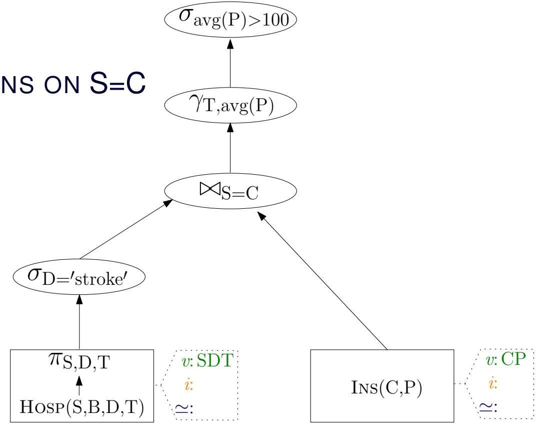
SELECT T, avg(P)
FROM Hosp JOIN Ins ON S=C
WHERE D='stroke'
GROUP BY T
HAVING avg(P)>100


$\sigma_{\text{avg(P)}>100}$
$\gamma_{\text{T,avg(P)}}$
$\bowtie_{\text{S=C}}$
$\sigma_{\text{D='stroke'}}$
$\pi_{\text{S,D,T}}$
Hosp(S,B,D,T)
Ins(C,P)

Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Plan with profiles

```
SELECT     T, avg(P)
FROM       HOSP JOIN INS ON S=C
WHERE      D=‘stroke’
GROUP BY   T
HAVING     avg(P)>100
```



$\sigma_{\text{avg(P)}>100}$

$\gamma_{\text{T,avg(P)}}$

$\bowtie_{\text{S=C}}$

$\sigma_{\text{D}='\text{stroke}'}$

$\pi_{\text{S,D,T}}$

Hosp(S,B,D,T)

v: SDT

i:

$\simeq$:

Ins(C,P)

v: CP

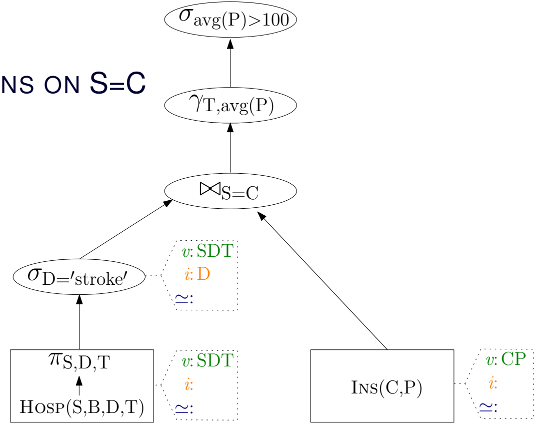
i:

$\simeq$:

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Plan with profiles

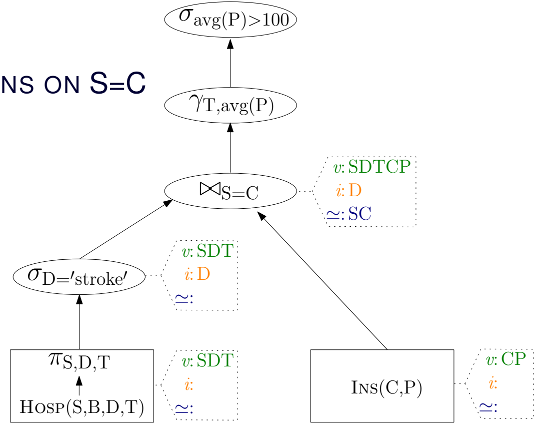
SELECT	T, avg(P)
FROM	HOSP JOIN INS ON S=C
WHERE	D=‘stroke’
GROUP BY	T
HAVING	avg(P)$>$100



HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Plan with profiles

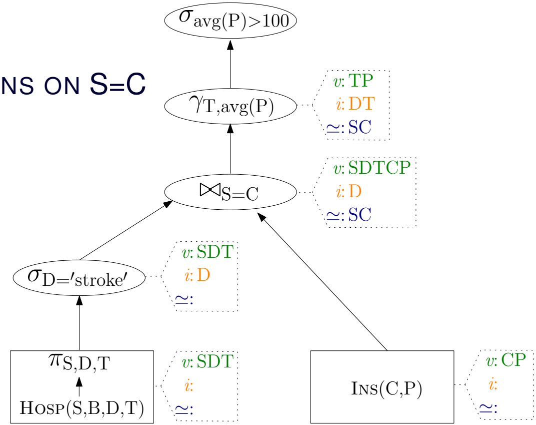
SELECT	T, avg(P)
FROM	HOSP JOIN INS ON S=C
WHERE	D=‘stroke’
GROUP BY	T
HAVING	avg(P)>100



HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Plan with profiles

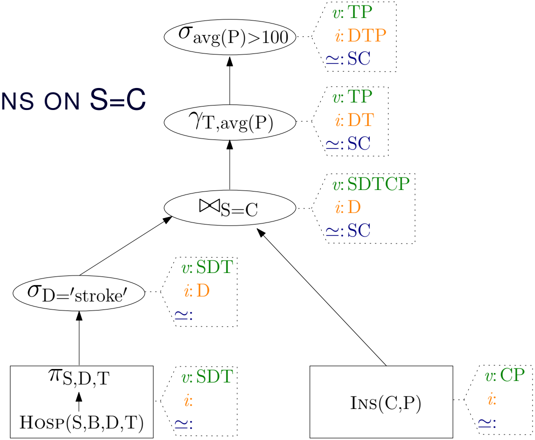
SELECT T, avg(P)
FROM HOSP JOIN INS ON S=C
WHERE D='stroke'
GROUP BY T
HAVING avg(P)>100



HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Plan with profiles

SELECT	T, avg(P)
FROM	HOSP JOIN INS ON S=C
WHERE	D='stroke'
GROUP BY	T
HAVING	avg(P)>100



HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Authorized visibility

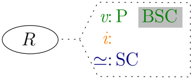
S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes

Authorized visibility

S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes



Relation

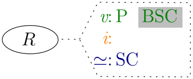
Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Authorized visibility

S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes



Relation

Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

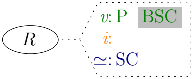
× cannot see P

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Authorized visibility

S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes



Relation

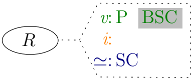
Subject	HOSP@ℍ	INS@𝕀	
ℍ	S B D T	C P	× cannot see P
𝕀	S B D T	C P	
𝕌	S D T	C P	× cannot see B (nor B)
𝕏	S D T	C P	× cannot see B (nor B)
𝕐	S B D T	C P	
ℤ	S D T	C P	× cannot see B (nor B)

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Authorized visibility

S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes



Relation

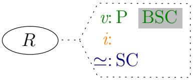
Subject	HOSP@ℍ				INS@𝕀		
ℍ	S	B	D	T	C	P	× cannot see P
𝕀	S	B	D	T	C	P	× no uniform vis. on ≃ { S ,C}
𝕌	S		D	T	C	P	× cannot see B (nor B)
𝕏	S		D	T	C	P	× cannot see B (nor B)
𝕐	S	B	D	T	C	P	
ℤ	S		D	T	C	P	× cannot see B (nor B)

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment)   INS(**C**ustomer, **P**remium)

Authorized visibility

S is authorized for R iff she has

- plaintext visibility on plaintext (visible or implicit) attributes
- plaintext or encrypted visibility on encrypted (visible or implicit) attributes
- uniform (plaintext or encrypted) visibility on equivalent attributes



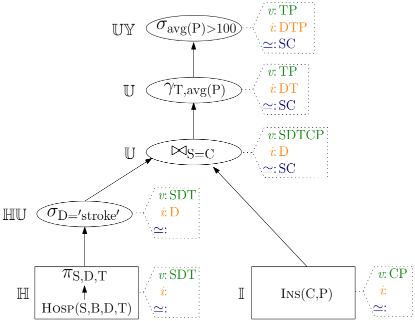
Relation

Subject	HOSP@ℍ				INS@𝕀		
ℍ	S	B	D	T	C	P	× cannot see P
𝕀	S	B	D	T	C	P	× no uniform vis. on ≃ {S,C}
𝕌	S		D	T	C	P	× cannot see B (nor B)
𝕏	S		D	T	C	P	× cannot see B (nor B)
𝕐	S	B	D	T	C	P	✓ authorized
ℤ	S		D	T	C	P	× cannot see B (nor B)

HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Compute assignments

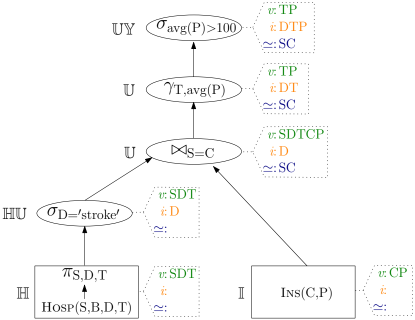
- Encrypting attributes not needed in plaintext for operands evaluation can increase candidates

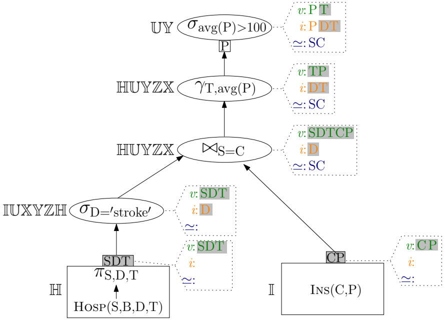


Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Compute assignments

- Encrypting attributes not needed in plaintext for operands evaluation can increase candidates





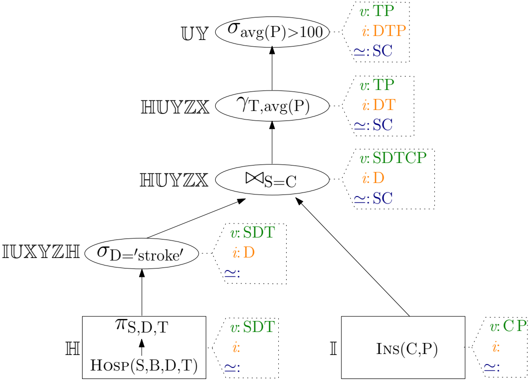
HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Relation

Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

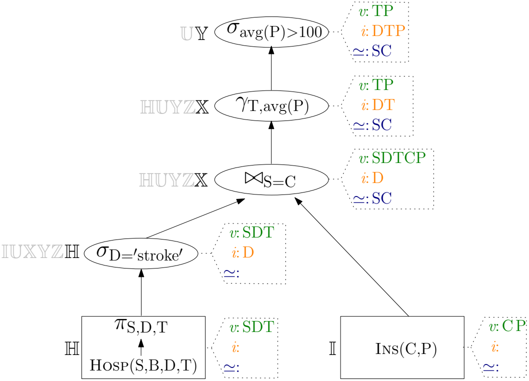


HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

	Relation					
Subject	Hosp@ℍ				Ins@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

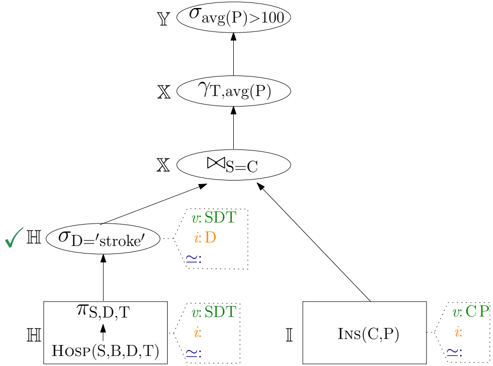


Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Subject		Relation	
		Hosp@ℍ	Ins@𝕀
	ℍ	S B D T	C P
	𝕀	S B D T	C P
	𝕌	S D T	C P
	𝕏	S D T	C P
	𝕐	S B D T	C P
	ℤ	S D T	C P



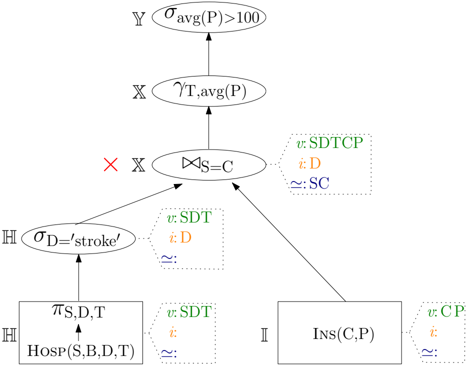
Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Relation

Subject	Hosp@ℍ	Ins@𝕀
ℍ	S B D T	C P
𝕀	S B D T	C P
𝕌	S D T	C P
𝕏	S D T	C P
𝕐	S B D T	C P
ℤ	S D T	C P



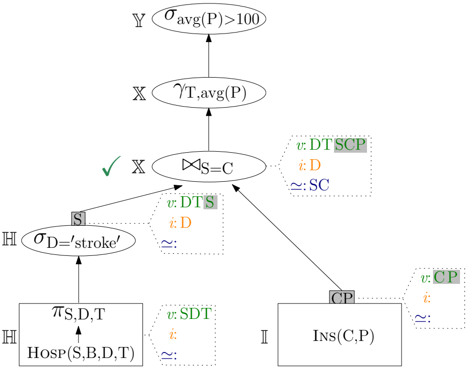
Hosp(**S**SN, **B**irth, **D**isease, **T**reatment)   Ins(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Relation

Subject	Hosp@ℍ				Ins@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P

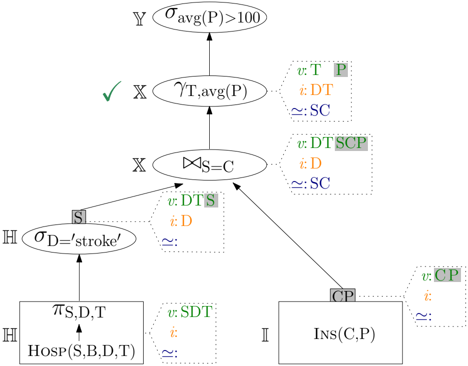


Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

	Relation					
Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P



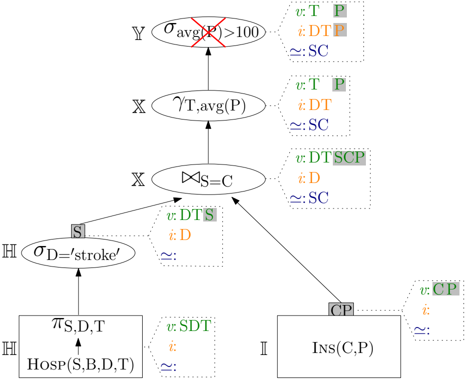
HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Relation

Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P



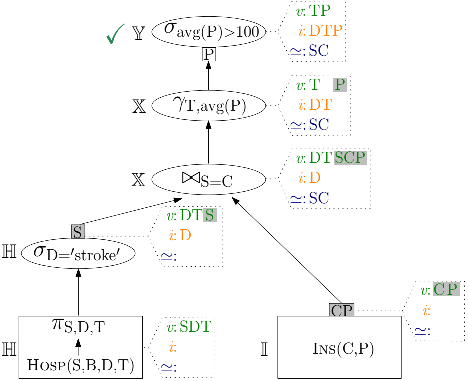
HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Minimally extended query plan

- Given a candidate for each node
  - encrypt attributes when needed for obeying authorizations
  - decrypt attributes when needed for the execution of an operation

Relation

Subject	HOSP@ℍ				INS@𝕀	
ℍ	S	B	D	T	C	P
𝕀	S	B	D	T	C	P
𝕌	S		D	T	C	P
𝕏	S		D	T	C	P
𝕐	S	B	D	T	C	P
ℤ	S		D	T	C	P



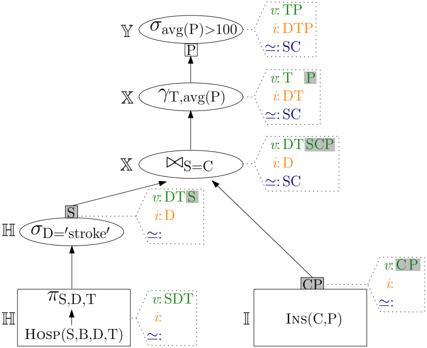
HOSP(**S**SN, **B**irth, **D**isease, **T**reatment) INS(**C**ustomer, **P**remium)

Key management

- Attributes in conditions comparing them must use the same key

  $\Longrightarrow$ attributes in the same equivalence set in the root use the same key

- Keys distributed to subjects in charge of enc/dec



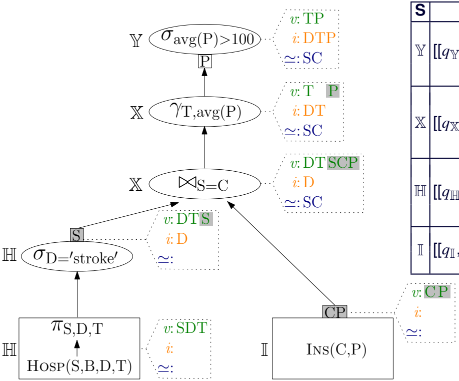
$\mathbb{Y}$ $\sigma_{\text{avg(P)}>100}$
v: TP
i: DTP
$\simeq$: SC
P
$\mathbb{X}$ $\gamma_{\text{T,avg(P)}}$
v: T P
i: DT
$\simeq$: SC
$\mathbb{X}$ $\bowtie_{\text{S=C}}$
v: DT SCP
i: D
$\simeq$: SC
S
$\mathbb{H}$ $\sigma_{\text{D='stroke'}}$
v: DT S
i: D
$\simeq$:
$\mathbb{H}$ $\pi_{\text{S,D,T}}$
HOSP(S,B,D,T)
v: SDT
i:
$\simeq$:
CP
$\mathbb{I}$ INS(C,P)
v: CP
i:
$\simeq$:

k_{SC}: same key for S and C given to $\mathbb{H}$ for encrypting S, $\mathbb{I}$ for encrypting C

k_{P}: key for P given to $\mathbb{I}$ for encryption, $\mathbb{Y}$ for decryption

Query dispatch

- Each sub-query is signed with the private key of the user and encrypted with the public key of the assignee



S	Receives (req$_S$)	Performs (q$_S$)

$\mathbb{Y}$	$[[q_{\mathbb{Y}},(P,k_P)]_{pri_{\mathbb{U}}}]_{pub_{\mathbb{Y}}}$	SELECT T,decrypt(P^k,k_P) AS P
		FROM ⟦req$_{\mathbb{X}}$⟧
		WHERE P >100

$\mathbb{X}$	$[[q_{\mathbb{X}},-]_{pri_{\mathbb{U}}}]_{pub_{\mathbb{X}}}$	SELECT T,avg(P^k) AS P^k
		FROM ⟦req$_{\mathbb{H}}$⟧ JOIN ⟦req$_{\mathbb{I}}$⟧ ON S^k=C^k
		GROUP BY T

$\mathbb{H}$	$[[q_{\mathbb{H}},(S,k_{SC})]_{pri_{\mathbb{U}}}]_{pub_{\mathbb{H}}}$	SELECT encrypt(S,k_{SC}),D,T
		FROM HOSP
		WHERE D='stroke'

$\mathbb{I}$	$[[q_{\mathbb{I}},(C,k_{SC})(P,k_P)]_{pri_{\mathbb{U}}}]_{pub_{\mathbb{I}}}$	SELECT encrypt(C,k_{SC}),encrypt(P,k_P)
		FROM INS

Summary

Novel and flexible approach for collaborative query evaluation

- authorities regulate access to their data
- users selectively involve external providers
- experiments show cost/performance savings in respect of authorizations

Several variations/open issues still need to be considered …

Other Considerations

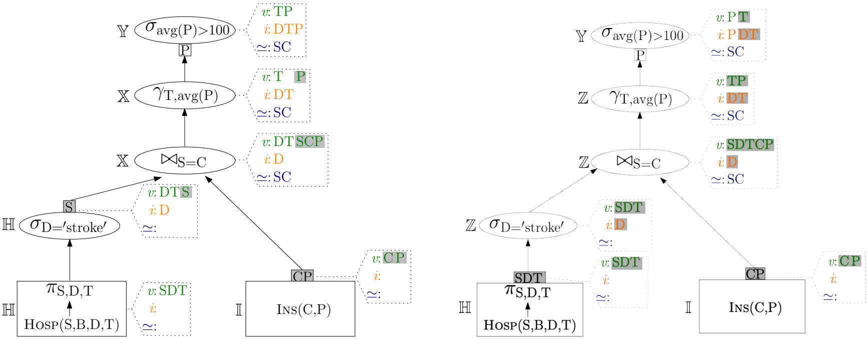
Economic/Performance Costs

S. De Capitani di Vimercati, S. Foresti, S. Jajodia, G. Livraga, S. Paraboschi, P. Samarati, "An Authorization Model for Query Execution in the Cloud," in *The VLDB Journal*, vol. 31, n. 3, May 2022, pp. 555-579.

Economic/performance costs

- Different authorized assignments may bear different economic/performance cost:
    - cost of encryption/decryption
    - cost of computation
    - cost of data transmission

Economic/performance costs – Example



⟹ determine an assignment that leverages on-the-fly encryption to minimize overall cost (including cost of encryption/decryption)

Hosp(**S**SN, **B**irth, **D**isease, **T**reatment) Ins(**C**ustomer, **P**remium)

Computing a minimum cost assignment

- Two steps approach:
  1. Compute candidates based on authorizations and assuming to encrypt all attributes not needed in plaintext for operands evaluation
  2. Determine an assignment such that the resulting query plan has minimum cost
- Minimization of the overall cost of query execution:

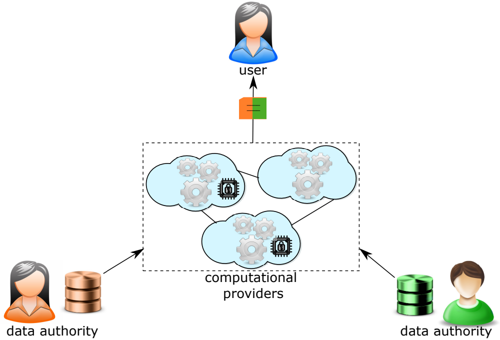
$$\min(\underbrace{\mathrm{OP\_EXEC}}_{\text{operation execution}} + \underbrace{\mathrm{ENC\_DEC}}_{\text{encryption/decryption}} + \underbrace{\mathrm{TRANSF}}_{\text{data transfer}})$$

Trusted Hardware

S. De Capitani di Vimercati, S. Foresti, S. Jajodia, G. Livraga, S. Paraboschi, P. Samarati, “Distributed Query Execution under Access Restrictions,” in *COSE*, vol. 127, April 2023

Trusted hardware – 1

- Providers could be equipped with trusted hardware components for query execution



⟹ need to integrate the use of a trusted hardware in the authorization model by properly defining its visibility over the data

Trusted hardware – 2

- Transmission of data to the trusted hardware is mediated by the subject hosting it

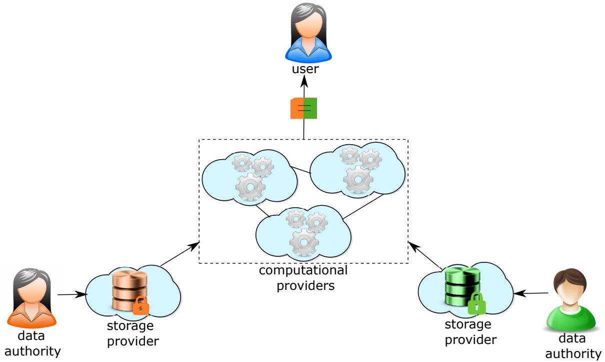
- Modeled as a different subject with authorizations more permissive than the ones of the subject hosting it
  - can access in plaintext at least the same attributes accessible to the hosting subject
  - can access in plaintext or encrypted a subset of the set of plaintext and encrypted attributes accessible to the hosting subject

Data Encryption in Storage

S. De Capitani di Vimercati, S. Foresti, S. Jajodia, G. Livraga, S. Paraboschi, P. Samarati, “Distributed Query Execution under Access Restrictions,” in *COSE*, vol. 127, April 2023

Encryption for protecting data in storage

Data stored at external storage providers might be encrypted by their owner for confidentiality



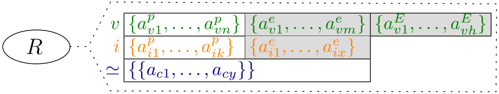
need mechanisms to support collaborative query execution over encrypted data

Collaborative computations over encrypted data

- In-storage encryption
  - is static and might not support the evaluation of the operations
  - is independently applied by each owner (different schemas and/or keys) and hence does not support comparison

  ⟹ re-encryption by authorized subjects to support collaborative query execution over data encrypted in storage

- Relation profile extended to capture the possible encrypted representation of attributes in storage



v	$\{a^p_{v1}, \ldots, a^p_{vn}\}$	$\{a^e_{v1}, \ldots, a^e_{vm}\}$	$\{a^E_{v1}, \ldots, a^E_{vh}\}$
i	$\{a^p_{i1}, \ldots, a^p_{ik}\}$	$\{a^e_{i1}, \ldots, a^e_{ix}\}$	
$\simeq$	$\{\{a_{c1}, \ldots, a_{cy}\}\}$		

Data and Computation Integrity

S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, R. Sassi, P. Samarati, “Sentinels and Twins: Effective Integrity Assessment for Distributed Computation,” in *IEEE TPDS*, vol. 34, n. 1, January 2023, pp. 108-122

Data and computation integrity – 1

- Data storage and processing may be performed by non trustworthy providers

- Need mechanisms that provide integrity for query results:
  - correctness: computed on genuine data
  - completeness: computed on the whole data collection
  - freshness: computed on the most recent version of the data

Data and computation integrity – 2

- Deterministic solutions based on a data structure (e.g., signature chains, Merkle hash trees, skip lists), need knowledge of the workload

- Probabilistic solutions based on dynamic insertion of control information:
  - markers/sentinels: fake tuples/tasks for which result is known
  - data job/replication: replicated tuples/tasks to check consistency in the result

Probabilistic approach for join queries

- A client, with the cooperation of the storage servers, can assess the integrity of joins performed by a computational cloud
- Protection techniques:
  - encryption makes data unintelligible
  - markers, fake tuples not recognizable as such by the computational cloud (and not colliding with real tuples)
  - twins, replication of existing tuples
- A marker missing or a twin appearing solo $\Longrightarrow$ integrity violation
- Probabilistic guarantee depending on the amount of control (markers and twins) inserted

On-the-fly encryption

- Server S encrypts $B(I, \mathit{Att})$, obtaining $B_k(I_k, B.\mathit{Tuple}_k)$
  - For each t in B, there is τ in B_k: $\tau[I_k]=E_k(t[I])$ and $\tau[B.\mathit{Tuple}_k]=E_k(t)$
  - E is a symmetric encryption function with key k
  - k is defined by the client and changes at every query
- Encryption provides data confidentiality

L

	I	Attr
l_1	a	Ann
l_2	b	Beth
l_3	c	Cloe

R

	I	Attr
r_1	a	flu
r_2	a	asthma
r_3	b	ulcer
r_4	e	hernia
r_5	e	flu
r_6	e	cancer

J

	L.I	L.Attr	R.I	R.Attr	
l_1	a	Ann	a	flu	r_1
l_1	a	Ann	a	asthma	r_2
l_2	b	Beth	b	ulcer	r_3

On-the-fly encryption

- Server S encrypts $B(I, Att)$, obtaining $B_k(I_k, B.Tuple_k)$
  - For each t in B, there is τ in B_k: $\tau[I_k]=E_k(t[I])$ and $\tau[B.Tuple_k]=E_k(t)$
  - E is a symmetric encryption function with key k
  - k is defined by the client and changes at every query
- Encryption provides data confidentiality

L_k

I_k	L.Tuple$_k$
α	λ_1
β	λ_2
γ	λ_3

R_k

I_k	R.Tuple$_k$
α	ρ_1
α	ρ_2
β	ρ_3
ε	ρ_4
ε	ρ_5
ε	ρ_6

J_k

L.I$_k$	L.Attr$_k$	R.I$_k$	R.Attr$_k$
α	λ_1	α	ρ_1
α	λ_1	α	ρ_2
β	λ_2	β	ρ_3

Markers

- Artificial tuples injected into L by S_l and R by S_r
  - not recognizable by the computational server
  - do not generate spurious tuples
  - inserted in a concerted manner to guarantee that they belong to the join result
- The absence of markers signals incompleteness of the join result

L

	I	Attr
l_1	a	Ann
l_2	b	Beth
l_3	c	Cloe

R

	I	Attr
r_1	a	flu
r_2	a	asthma
r_3	b	ulcer
r_4	e	hernia
r_5	e	flu
r_6	e	cancer

J

	L.I	L.Attr	R.I	R.Attr	
l_1	a	Ann	a	flu	r_1
l_1	a	Ann	a	asthma	r_2
l_2	b	Beth	b	ulcer	r_3

Markers

- Artificial tuples injected into L by S_l and R by S_r
  - not recognizable by the computational server
  - do not generate spurious tuples
  - inserted in a concerted manner to guarantee that they belong to the join result
- The absence of markers signals incompleteness of the join result

L^*

	I	Attr
l_1	a	Ann
l_2	b	Beth
l_3	c	Cloe
m_1	x	*marker*$_1$

R^*

	I	Attr
r_1	a	flu
r_2	a	asthma
r_3	b	ulcer
r_4	e	hernia
r_5	e	flu
r_6	e	cancer
m_2	x	*marker*$_2$

J^*

	L.I	L.Attr	R.I	R.Attr	
l_1	a	Ann	a	flu	r_1
l_1	a	Ann	a	asthma	r_2
l_2	b	Beth	b	ulcer	r_3
m_1	x	*marker*$_1$	x	*marker*$_2$	m_2

Twins

- Duplicates of tuples that satisfy condition C_{twin} that
  - is defined on the join attribute I
  - tunes the percentage p_t of twins
  - is defined by the client and communicated to S_l and S_r
- Twin pairs are not recognizable by the computational server
- A twin appearing solo signals incompleteness of the join result

L

	I	**Attr**
l_1	a	Ann
l_2	b	Beth
l_3	c	Cloe

R

	I	**Attr**
r_1	a	flu
r_2	a	asthma
r_3	b	ulcer
r_4	e	hernia
r_5	e	flu
r_6	e	cancer

J

	L.I	**L.Attr**	**R.I**	**R.Attr**	
l_1	a	Ann	a	flu	r_1
l_1	a	Ann	a	asthma	r_2
l_2	b	Beth	b	ulcer	r_3

Twins

- Duplicates of tuples that satisfy condition C_{twin} that
  - is defined on the join attribute I
  - tunes the percentage p_t of twins
  - is defined by the client and communicated to S_l and S_r
- Twin pairs are not recognizable by the computational server
- A twin appearing solo signals incompleteness of the join result

L^*

	I	Attr
l_1	a	Ann
l_2	b	Beth
l_3	c	Cloe
$\bar{l}_2$	$\bar{\text{b}}$	Beth

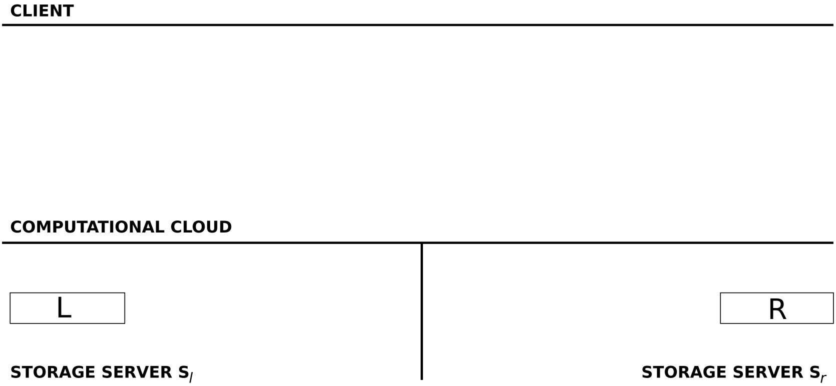
R^*

	I	Attr
r_1	a	flu
r_2	a	asthma
r_3	b	ulcer
r_4	e	hernia
r_5	e	flu
r_6	e	cancer
$\bar{r}_3$	$\bar{\text{b}}$	ulcer

J^*

	L.I	L.Attr	R.I	R.Attr	
l_1	a	Ann	a	flu	r_1
l_1	a	Ann	a	asthma	r_2
l_2	b	Beth	b	ulcer	r_3
$\bar{l}_2$	$\bar{\text{b}}$	Beth	$\bar{\text{b}}$	ulcer	$\bar{r}_3$

Probabilistic approach for join queries – Example



CLIENT

COMPUTATIONAL CLOUD

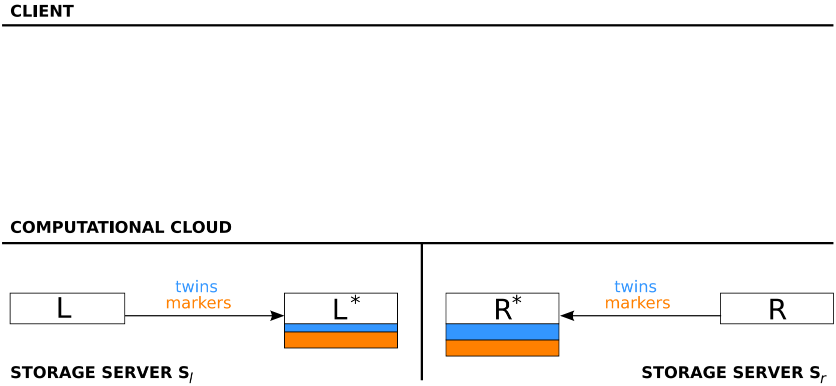
L

R

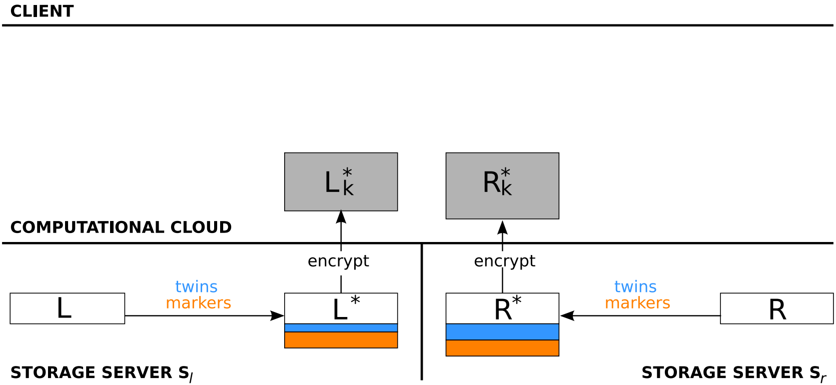
STORAGE SERVER S_l

STORAGE SERVER S_r

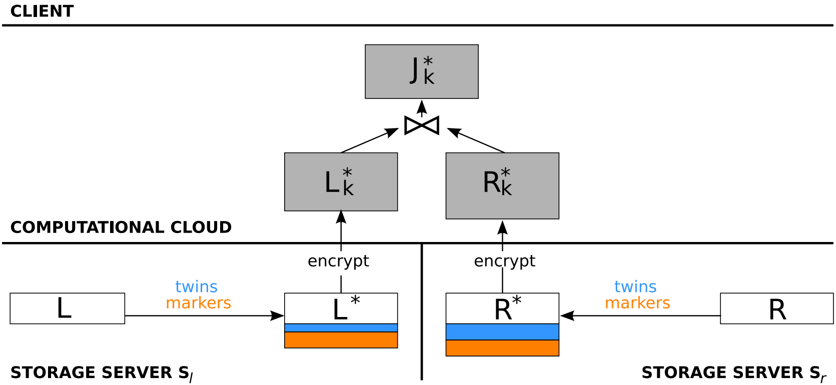
Probabilistic approach for join queries – Example



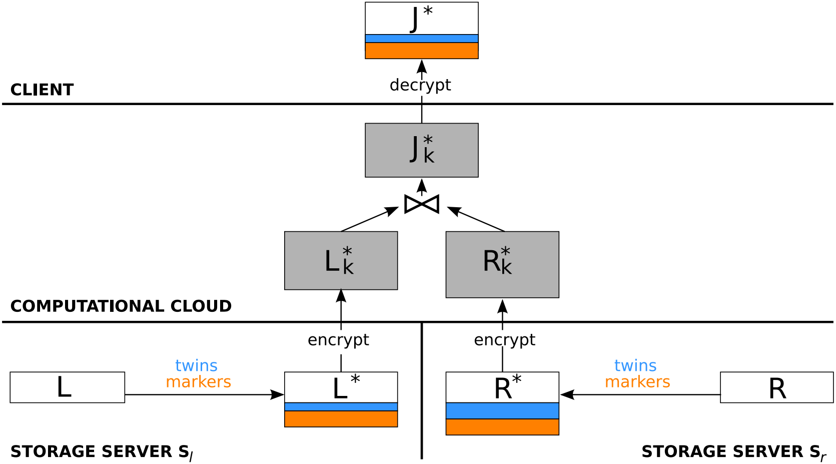
Probabilistic approach for join queries – Example



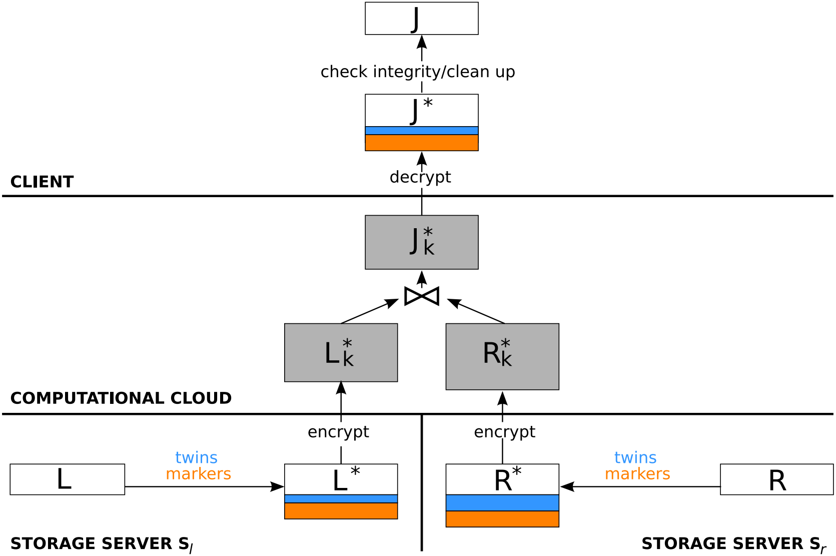
Probabilistic approach for join queries – Example



Probabilistic approach for join queries – Example



Probabilistic approach for join queries – Example



Markers and twins: Integrity guarantees

- The guarantee offered by markers and twins can be measured as the probability of the computational cloud to go undetected when omitting tuples

- Markers and twins offer complementary protection:
  - Twins are twice as effective as markers, but loose their effectiveness when the computational cloud omits a large fraction of tuples (extreme case: all tuples omitted)
  - Markers allow detecting extreme behavior (all tuples omitted) and provide effective when the computational cloud omits a large fraction of tuples

Markers and twins: Some considerations

- For 1:n joins, join profile needs to be protected (salts and buckets)

- Markers and twins need to be non recognizable

- Consideration of generic computations involving different sets of workers

Conclusions

- Advancements in ICT and networks:
  - enable new and better applications and services, bringing social and economic benefits
  - need to address new security and privacy risks and challenges

… towards allowing society to fully benefit from information technology while enjoying security and privacy